Chapter 12

Stochastic Optimal Control

In previous chapters we assumed that the state variables of the system are known with certainty. When the variables are outcomes of a random phenomenon, the state of the system is modeled as a stochastic process. Specifically, we now face 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 Sect. 12.1, 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 Sect. 12.5.

In Sect. 12.2, we will extend the production planning model of Chap. 6 to allow for some uncertain disturbances. We will obtain an optimal production policy for the stochastic production planning problem thus formulated. In Sect. 12.3, we will solve an optimal stochastic advertising

S. P. Sethi, *Optimal Control Theory*,
https://doi.org/10.1007/978-3-319-98237-3_12

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 Sect. 7.2.4. In Sect. 12.4, 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 Sect. 12.5, we will conclude the chapter by mentioning other types of stochastic optimal control problems that arise in practice.

12.1 Stochastic Optimal Control

In Appendix D.1 on the Kalman filter, we obtain the optimal state estimation for linear systems with noise and noisy measurements. In Sect. 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 Eq. (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 uses dynamic programming or the Hamilton-Jacobi-Bellman framework rather than a stochastic maximum principle. 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

$$E\left[\int_0^T F(X_t, U_t, t)dt + S(X_T, T)\right], \tag{12.1}$$

where X_t is the state at time t and U_t is the control at time t, and together they are required to satisfy the Itô stochastic differential equation

$$dX_t = f(X_t, U_t, t)dt + G(X_t, U_t, t)dZ_t, \quad X_0 = x_0, \tag{12.2}$$

where $Z_t,\ t \in [0, T]$ is a standard Wiener process.

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 (12.2) 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 (12.2) is a scalar equation, the subscript t here represents only time t. Thus, writing X_t, U_t, and Z_t in place of writing $X(t), U(t)$, and $Z(t)$, respectively, will not cause any confusion and, at the same time, will eliminate the need for writing many parentheses.

To solve the problem defined by (12.1) and (12.2), let $V(x, t)$, known as the *value function*, be the expected value of the objective function (12.1) from t to T, when an optimal policy is followed from t to T, given $X_t = x$. Then, by the principle of optimality,

$$V(x, t) = \max_U E[F(x, U, t)dt + V(x + dX_t, t + dt)]. \tag{12.3}$$

By Taylor's expansion, we have

$$\begin{aligned} V(x + dX_t, t + dt) = V(x, t) \quad & +V_t dt + V_x dX_t + \tfrac{1}{2}V_{xx}(dX_t)^2 \\ & +\tfrac{1}{2}V_{tt}(dt)^2 + \tfrac{1}{2}V_{xt}dX_t dt \\ & +\text{higher-order terms.} \end{aligned} \tag{12.4}$$

From (12.2), we can formally write

$$(dX_t)^2 = f^2(dt)^2 + G^2(dZ_t)^2 + 2fGdZ_t dt, \tag{12.5}$$
$$dX_t dt = f(dt)^2 + GdZ_t dt. \tag{12.6}$$

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 stochastic calculus:

$$(dZ_t)^2 = dt,\ dZ_t dt = 0,\ dt^2 = 0. \tag{12.7}$$

Substitute (12.4) into (12.3) and use (12.5), (12.6), (12.7), and the property that $E[dZ_t] = 0$ to obtain

$$V = \max_U \left[F dt + V + V_t dt + V_x f dt + \frac{1}{2} V_{xx} G^2 dt + o(dt) \right]. \qquad (12.8)$$

Note that we have suppressed the arguments of the functions involved in (12.8).

Canceling the term V on both sides of (12.8), dividing the remainder by dt, and letting $dt \to 0$, we obtain the Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \max_U [F + V_t + V_x f + \frac{1}{2} V_{xx} G^2] \qquad (12.9)$$

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

$$V(x, T) = S(x, T). \qquad (12.10)$$

Just as we had introduced a current-value formulation of the maximum principle in Sect. 3.3, let us derive a current-value version of the HJB equation here. For this, in a way similar to (3.29), we write the objective function to be maximized as

$$E \int_0^T [\phi(X_t, U_t) e^{-\rho t} + \psi(X_T) e^{-\rho T}]. \qquad (12.11)$$

We can relate this to (12.1) by setting

$$F(X_t, U_t, t) = \phi(X_t, U_t) e^{-\rho t} \text{ and } S(X_T, T) = \psi(X_T) e^{-\rho T}. \qquad (12.12)$$

It is important to mention that the explicit dependence on time t in (12.11) is only via the discounting term. If it were not the case, there would be no advantage in formulating the current-value version of the HJB equation.

Rather than develop the current-value HJB equation in a manner of developing (12.9), we will derive it from (12.9) itself. For this we define the current-valued value function

$$\tilde{V}(x, t) = V(x, t) e^{\rho t}. \qquad (12.13)$$

Then we have

$$V_t = \tilde{V}_t e^{-\rho t} - \rho \tilde{V} e^{-\rho t}, \; V_x = \tilde{V}_x e^{-\rho t} \text{ and } V_{xx} = \tilde{V}_{xx} e^{-\rho t}. \qquad (12.14)$$

By using these and (12.12) in (12.9), we obtain

$$0 = \max_U [\phi e^{-\rho t} + \tilde{V} e^{-\rho t} - \rho \tilde{V} e^{-\rho t} + V_x f e^{-\rho t} + \frac{1}{2} V_{xx} G^2 e^{-\rho t}].$$

Multiplying by $e^{\rho t}$ and rearranging terms, we get

$$\rho \tilde{V} = \max_U [\phi + \tilde{V}_t + \tilde{V}_x f + \frac{1}{2} \tilde{V}_{xx} G^2]. \tag{12.15}$$

Moreover, from (12.12), (12.13), and (12.10), we can get the boundary condition

$$\tilde{V}(x, T) = \psi(x). \tag{12.16}$$

Thus, we have obtained (12.15) and (12.16) as the current-value HJB equation.

To obtain its infinite-horizon version, it is generally the case that we remove the explicit dependence on t from the function f and G in (12.2), and also assume that $\psi \equiv 0$. With that, the dynamics (12.2) and the objective function (12.11) change, respectively, to

$$dX_t = f(X_t, U_t)dt + G(X_t, U_t)dZ_t, \; X_0 = x_0, \tag{12.17}$$

$$E \int_0^\infty \phi(X_t, U_t) e^{-\rho t} dt. \tag{12.18}$$

It should then be obvious that $\tilde{V}_t = 0$, and we can obtain the infinite-horizon version of (12.15) as

$$\rho \tilde{V} = \max_U [\phi + \tilde{V}_x f + \frac{1}{2} \tilde{V}_{xx} G^2]. \tag{12.19}$$

As for its boundary condition, (12.16) is replaced by a growth condition that is the same, in general, as the growth of the function $\phi(x, U)$ in x. For example, if $\phi(x, U)$ is quadratic in x, we would look for a value function $\tilde{V}(x)$ to be of quadratic growth. See Beyer et al. (2010), Chapter 3, for a related discussion of a polynomial growth case in the discrete time setting.

If we can find a solution of the HJB equation with the given boundary condition (or an appropriate growth condition in the infinite horizon case), then a result called a *verification theorem* suggests that we can construct an optimal feedback control $U^*(x, t)$ (or $U^*(x)$ in the infinite horizon case) by maximizing the right-hand side of the HJB equation with respect U. For further details and extension when the value function is not smooth enough and thus not a classical solution of the HJB equation, see Fleming and Rishel (1975), Yong and Zhou (1999), and Fleming and Soner (1992).

In the next three sections, we will apply this procedure to solve problems in production, marketing and finance.

12.2 A Stochastic Production Inventory Model

In Sect. 6.1.1, we 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:

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

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

$$dI_t = (P_t - S)dt + \sigma dZ_t, \quad I_0 \text{ given}, \tag{12.20}$$

where I_0 denotes the initial inventory level. As mentioned in Appendix Sect. D.2, 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:

$$\max E\left\{BI_T - \int_0^T (P_t^2 + I_t^2)dt\right\}. \tag{12.21}$$

It can be interpreted as maximization of the terminal salvage value less the cost of production and inventory assumed to be quadratic. In Exercise 12.1, 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 Sect. 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 Sethi and Thompson (1981a) 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

$$0 = \max_P[-(P^2 + x^2) + V_t + V_x(P - S) + \frac{1}{2}\sigma^2 V_{xx}] \tag{12.22}$$

with the boundary condition

$$V(x,T) = Bx. \tag{12.23}$$

Note that these are applications of (12.9) and (12.10) to the production planning problem.

It is now possible to maximize the expression inside the bracket of (12.22) with respect to P by taking its derivative with respect to P and setting it to zero. This procedure yields

$$P^*(x,t) = \frac{V_x(x,t)}{2}. \tag{12.24}$$

Substituting (12.24) into (12.22) yields the equation

$$0 = \frac{V_x^2}{4} - x^2 + V_t - SV_x + \frac{1}{2}\sigma^2 V_{xx}, \tag{12.25}$$

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 (12.23). The solution of (12.25) is considered in the next section.

Remark 12.1 It is important to remark that if the production rate were restricted to be nonnegative, then, as in Remark 6.1, (12.24) would be changed to

$$P^*(x,t) = \max\left[0, \frac{V_x(x,t)}{2}\right]. \tag{12.26}$$

Substituting (12.26) into (12.23) would give us a partial differential equation which must be solved numerically. We will not consider (12.26) further in this chapter.

12.2.1 Solution for the Production Planning Problem

To solve Eq. (12.25) with the boundary condition (12.23) we let

$$V(x,t) = Q(t)x^2 + R(t)x + M(t). \tag{12.27}$$

Then,

$$\begin{aligned} V_t &= \dot{Q}x^2 + \dot{R}x + \dot{M}, & (12.28)\\ V_x &= 2Qx + R, & (12.29)\\ V_{xx} &= 2Q, & (12.30) \end{aligned}$$

where $\dot{Y}$ denotes dY/dt. Substituting (12.28)–(12.30) in (12.25) and collecting terms gives

$$x^2[\dot{Q} + Q^2 - 1] + x[\dot{R} + RQ - 2SQ] + \dot{M} + \frac{R^2}{2} - RS + \sigma^2 Q = 0. \tag{12.31}$$

Since (12.31) must hold for any value of x, we must have

$$\begin{aligned} \dot{Q} &= 1 - Q^2, \;\; Q(T) = 0, & (12.32)\\ \dot{R} &= 2SQ - RQ, \;\; R(T) = B, & (12.33)\\ \dot{M} &= RS - \frac{R^2}{4} - \sigma^2 Q, \;\; M(T) = 0, & (12.34) \end{aligned}$$

where the boundary conditions for the system of simultaneous differential equations (12.32), (12.33), and (12.34) are obtained by comparing (12.27) with the boundary condition $V(x,T) = Bx$ of (12.23).

To solve (12.32), 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

$$Q = \frac{y-1}{y+1}, \tag{12.35}$$

where

$$y = e^{2(t-T)}. \tag{12.36}$$

Since S is assumed to be a constant, we can reduce (12.33) 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

$$R = 2S + \frac{2(B-2S)\sqrt{y}}{y+1}. \tag{12.37}$$

Having obtained solutions for R and Q, we can easily express (12.34) as

$$M(t) = -\int_t^T [R(\tau)S - (R(\tau))^2/4 - \sigma^2 Q(\tau)]d\tau. \tag{12.38}$$

The optimal control is defined by (12.24), and the use of (12.35) and (12.37) yields

$$P^*(x,t) = V_x/2 = Qx + R/2 = S + \frac{(y-1)x + (B-2S)\sqrt{y}}{y+1}. \tag{12.39}$$

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

$$P_t^* = P^*(I_t^*, t) = S + \frac{(e^{2(t-T)} - 1)I_t^* + (B-2S)e^{(t-T)}}{e^{2(t-T)} + 1}, \tag{12.40}$$

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 (12.40).

Remark 12.2 The optimal production rate in (12.39) 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 12.3 If the demand rate S were time-dependent, it would have changed the solution of (12.33). Having computed this new solution in place of (12.37), we can once again obtain the optimal control as $P^*(x,t) = Qx + R/2$.

Remark 12.4 Note that when $T \to \infty$, we have $y \to 0$ and

$$P^*(x,t) \to S - x, \tag{12.41}$$

but the undiscounted objective function value (12.21) 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 12.2.

Remark 12.5 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 Fig. 12.1. 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.

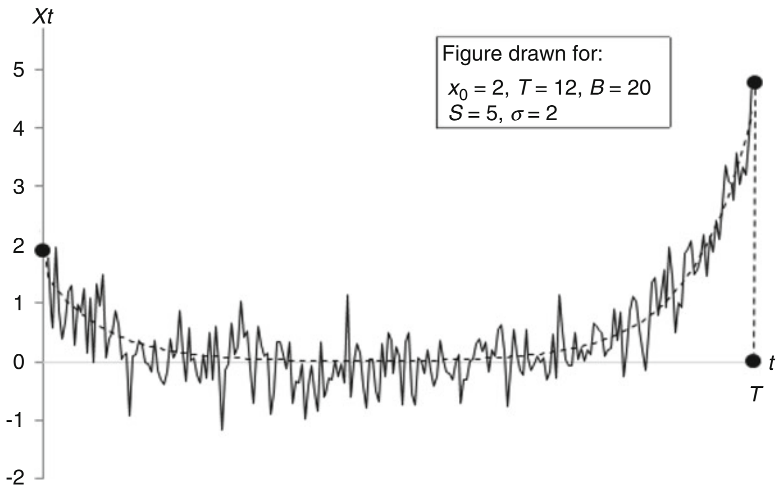

Figure 12.1: A sample path of optimal production rate I_t^* with $I_0 = x_0 > 0$ and $B > 0$

12.3 The Sethi Advertising Model

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

$$\begin{cases} \max\ E\left[\int_0^\infty e^{-\rho t}(\pi X_t - U_t^2)dt\right] \\ \text{subject to} \\ dX_t = (rU_t\sqrt{1-X_t} - \delta X_t)dt + \sigma(X_t)dZ_t,\ X_0 = x_0, \\ U_t \geq 0, \end{cases} \tag{12.42}$$

where X_t is the market share and U_t is the rate of advertising at time t, and where, as specified in Sect. 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 integral represents the total value of the discounted profit stream on a sample path. The objective in (12.42) is, therefore, to maximize the expected value of the total discounted profit.

The Sethi 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 (12.42) 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, the linear cost of advertising U in (7.43) is replaced by a quadratic cost of advertising U_t^2 in (12.42). The control constraint $0 \leq U \leq Q$ in (7.43) is replaced by simply $U_t \geq 0$. The addition of the diffusion term yields a stochastic optimal control problem as expressed in (12.42).

An important consideration in choosing the function $\sigma(x)$ should be that the solution X_t to the Itô equation in (12.42) remains inside the interval $[0, 1]$. Merely requiring that the initial condition $x_0 \in [0, 1]$, as in Sect. 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 (12.42):

$$\sigma(x) > 0,\ x \in (0,1) \text{ and } \sigma(0) = \sigma(1) = 0. \tag{12.43}$$

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

$$U(x) \geq 0,\ x \in (0,1],\ \text{ and } U(0) > 0, \tag{12.44}$$

the Itô equation in (12.42) 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 (12.44), 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 that 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 Sect. 12.1, we can write the HJB equation as

$$\rho V(x) = \max_U \left[\pi x - U^2 + V_x(rU\sqrt{1-x} - \delta x) + V_{xx}(\sigma(x))^2/2\right]. \tag{12.45}$$

Maximization of the RHS of (12.45) can be accomplished by taking its derivative with respect to U and setting it to zero. This gives

$$U^*(x) = \frac{rV_x\sqrt{1-x}}{2}. \tag{12.46}$$

Substituting of (12.46) in (12.45) and simplifying the resulting expression yields the HJB equation

$$\rho V(x) = \pi x + \frac{V_x^2 r^2(1-x)}{4} - V_x\delta x + \frac{1}{2}\sigma^2(x)V_{xx}. \tag{12.47}$$

As shown in Sethi (1983b), a solution of (12.47) is

$$V(x) = \bar{\lambda}x + \frac{\bar{\lambda}^2 r^2}{4\rho}, \tag{12.48}$$

where

$$\bar{\lambda} = \frac{\sqrt{(\rho+\delta)^2 + r^2\pi} - (\rho+\delta)}{r^2/2}, \tag{12.49}$$

as derived in Exercise 7.37. In Exercise 12.3, you are asked verify that (12.48) and (12.49) solve the HJB equation (12.47).

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

$$U^*(x) = \frac{r\bar{\lambda}\sqrt{1-x}}{2}. \tag{12.50}$$

Note that $U^*(x)$ satisfies the conditions in (12.44).

As in Exercise 7.37, it is easy to characterize (12.50) as

$$U_t^* = U^*(X_t) = \begin{cases} > \bar{U} & \text{if } X_t < \bar{X}, \\ = \bar{U} & \text{if } X_t = \bar{X}, \\ < \bar{U} & \text{if } X_t > \bar{X}, \end{cases} \tag{12.51}$$

where

$$\bar{X} = \frac{r^2\bar{\lambda}/2}{r^2\bar{\lambda}/2+\delta} \tag{12.52}$$

and

$$\bar{U} = \frac{r\bar{\lambda}\sqrt{1-\bar{x}}}{2}, \tag{12.53}$$

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 (12.42). 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.

12.4 An Optimal Consumption-Investment Problem

In Example 1.3 in Chap. 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 the possibility of investing a part of his wealth in a risky security or stock that earns an expected rate of return that equals $\alpha > r$. Rich, now known as Rich Investor, must 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 the initial deposit 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

$$dS_t = rS_t dt,\ S_0 \geq 0. \tag{12.54}$$

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

$$\frac{dP_t}{P_t} = \alpha dt + \sigma dZ_t,\ P_0 > 0, \tag{12.55}$$

or simply,

$$dP_t = \alpha P_t dt + \sigma P_t dZ_t,\ P_0 > 0, \tag{12.56}$$

where $P_0 > 0$ is the given initial stock price, 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 12.6 The LHS in (12.55) 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 (12.55) is often referred to as a *logarithmic Brownian Motion.* It is important to note from (12.56) that P_t remains nonnegative at any $t > 0$ on account of the fact that the price process has almost surely continuous sample paths (see Sect. D.2). This property nicely captures the limited liability that is incurred in owning a share of stock.

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

W_t = the wealth at time t,

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

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 rigorously obtain 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

$$
\begin{aligned}
dW_t &= Q_t W_t \alpha dt + Q_t W_t \sigma dZ_t + (1 - Q_t) W_t r dt - C_t dt \\
&= (\alpha - r) Q_t W_t dt + (r W_t - C_t) dt + \sigma Q_t W_t dZ_t, \ W_0 \text{ given,}
\end{aligned}
\tag{12.57}
$$

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, $C_t dt$ represents the amount of consumption during the interval from t to $t + dt$.

In deriving (12.57), 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 to consumption as well as the change in share price. For a rigorous development of (12.57) from (12.54) and (12.55), see Harrison and Pliska (1981).

Since Rich can borrow an unlimited amount 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 defined as

$$T = \inf\{t \geq 0 | W_t = 0\}. \tag{12.58}$$

This special type of random variable is 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:

$$\max \left\{ J = E\left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\}, \tag{12.59}$$

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:

$$\begin{cases} \max \left\{ J = E\left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\} \\ \text{subject to} \\ dW_t = (\alpha - r) Q_t W_t dt + (r W_t - C_t) dt + \sigma Q_t W_t dZ_t, \; W_0 \text{ given}, \\ C_t \geq 0. \end{cases} \tag{12.60}$$

As in the infinite horizon problem of Sect. 12.2, 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 Sect. 12.1, the HJB equation satisfied by the value function $V(x)$ for problem (12.60) can be written as

$$\begin{cases} \rho V(x) = \max\limits_{C \geq 0, Q} \quad [(\alpha - r) Q x V_x + (rx - C) V_x \\ \qquad\qquad\qquad +(1/2) Q^2 \sigma^2 x^2 V_{xx} + U(C)], \\ V(0) = B. \end{cases} \tag{12.61}$$

This problem and a number of its generalizations are solved explicitly in Sethi (1997a). Here we shall confine ourselves in solving a simpler problem resulting from the following considerations.

It is shown in Karatzas et al. (1986), reproduced as Chapter 2 in Sethi (1997a), that when $B \leq U(0)/\rho$, no bankruptcy will occur. This should be intuitively obvious because if Rich goes bankrupt at any time $T > 0$, he receives B at that time, whereas by not going bankrupt at that time he reaps the utility of strictly more than $U(0)/\rho$ on account of consumption from time T onward. It is shown furthermore that if $U'(0) = \infty$, then the optimal consumption rate will be strictly positive. This is because even an infinitesimally small positive consumption rate results in a proportionally large amount of utility on account of the infinite marginal utility at zero consumption level. A popular utility function used in the literature is

$$U(C) = \ln C, \tag{12.62}$$

which was also used in Example 1.3. This function gives an infinite marginal utility at zero consumption, i.e.,

$$U'(0) = 1/C|_{C=0} = \infty. \tag{12.63}$$

We also assume $B = U(0)/\rho = -\infty$. These assumptions imply 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 (12.61) with respect to Q and C and equating the resulting expressions to zero. Thus,

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

i.e.,

$$Q^*(x) = -\frac{(\alpha - r)V_x}{x\sigma^2 V_{xx}}, \tag{12.64}$$

and

$$C^*(x) = \frac{1}{V_x}. \tag{12.65}$$

Substituting (12.64) and (12.65) in (12.61) allows us to remove the max operator from (12.61), and provides us with the equation

$$\rho V(x) = -\frac{\gamma (V_x)^2}{V_{xx}} + \left(rx - \frac{1}{V_x}\right) V_x - \ln V_x, \tag{12.66}$$

where

$$\gamma = \frac{(\alpha - r)^2}{2\sigma^2}. \tag{12.67}$$

This is a nonlinear ordinary differential equation that appears to be quite difficult to solve. However, Karatzas et al. (1986) used a change of variable that transforms (12.66) into a second-order, linear, ordinary differential equation, which has a known solution. For our purposes, we will simply guess that the value function is of the form

$$V(x) = A \ln x + B, \tag{12.68}$$

where A and B are constants, and obtain the values of A and B by substitution in (12.66). Using (12.68) in (12.66), 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 $B = (r - \rho + \gamma)/\rho^2 + \ln \rho/\rho$. By substituting these values in (12.68), we obtain

$$V(x) = \frac{1}{\rho}\ln(\rho x) + \frac{r - \rho + \gamma}{\rho^2},\ x \geq 0. \tag{12.69}$$

In Exercise 12.4, you are asked by a direct substitution in (12.66) to verify that (12.69) is indeed a solution of (12.66). Moreover, $V(x)$ defined in (12.69) is strictly concave, so that our concavity assumption made earlier is justified.

From (12.69), it is easy to show that (12.64) and (12.65) yield the following feedback policies:

$$Q^*(x) = \frac{\alpha - r}{\sigma^2}, \tag{12.70}$$

$$C^*(x) = \rho x. \tag{12.71}$$

The investment policy (12.70) says that the optimal fraction of the wealth invested in the risky stock is $(\alpha - r)/\sigma^2$, i.e.,

$$Q_t^* = Q^*(W_t) = \frac{\alpha - r}{\sigma^2},\ t \geq 0, \tag{12.72}$$

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

$$C_t^* = C^*(W_t) = \rho W_t,\ t \geq 0. \tag{12.73}$$

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

12.5 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, 2018), Elliott et al. (1995), and Bensoussan et al. (2010).

For applications of stochastic optimal control to manufacturing problems, see Sethi and Zhang (1994a), Yin and Zhang (1997), Sethi et al. (2005), Bensoussan (2011), and Bensoussan et al. (2007b,c,d, 2008a,b, 2009a,b,c). For applications to problems in finance, see Sethi (1997a), Karatzas and Shreve (1998), and Bensoussan et al. (2009d). For applications in marketing, see Tapiero (1988), Raman (1990), and Sethi and Zhang (1995b). For applications of stochastic optimal control to economics including economics of natural resources, see, e.g., Pindyck (1978a,b), 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 12

E 12.1 Solve the production-inventory problem with the state equation (12.20) 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 12.2 Formulate and solve the discounted infinite-horizon version of the stochastic production planning model of Sect. 12.2. Specifically, assume $B = 0$ and replace the objective function in (12.21) by

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

E 12.3 Verify by direct substitution that the value function defined by (12.48) and (12.49) solves the HJB equation (12.47).

E 12.4 Verify by direct substitution that the value function in (12.69) solves the HJB equation (12.66).

E 12.5 Solve the consumption-investment problem (12.60) with the utility function $U(C) = C^\beta$, $0 < \beta < 1$, and $B = 0$.

E 12.6 Solve Exercise 12.5 when $U(C) = -C^\beta$ with $\beta < 0$ and $B = -\infty$.

E 12.7 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

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

Here $s > 0$ denotes a minimal subsistence consumption, and we assume $0 < \rho < 1$. 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 12.8 Solve the consumption-investment problem:

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

subject to

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

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.