



CHAPTER 13

STOCHASTIC OPTIMAL CONTROL

- The state of the system is represented by a controlled *stochastic process*.
- A decision maker is faced with the problem of making good estimates of these state variables from noisy measurements on functions of them.
- The process of estimating the values of the state variables is called *optimal filtering*.

13.1 THE KALMAN FILTER

There are two types of random disturbances present.

- *Measurement noise*: This type of disturbances arise because of imprecise measurement instruments, inaccurate recording systems, etc. In many cases the measurement technique involves observations of functions of state variables, from which the values of some or all of the state variables are inferred.
- *System noise*: The system itself is subjected to random disturbances. For instance, sales may follow a stochastic process, which affects the system equation (6.1) relating inventory, production, and sales. In the cash balance example, the demand for cash as well as the interest rates in (5.1) and (5.2) can be represented by stochastic processes.

In analyzing systems, in which one or both of these kinds of noises are present, it is important to be able to make good *estimates* of the values of the state variables. In particular, many optimal control problems require such estimates in order to determine the optimal controls; see Appendix D.1.1. The process of estimating the current values of the state variables given the past measurements is called *filtering*; see Kalman (1960a, 1960b), Sorenson (1966), and Bryson and Ho (1969).

A DYNAMIC STOCHASTIC SYSTEM

Consider a dynamic stochastic system in discrete time described by the difference equation

$$x^{t+1} - x^t = A_t x^t + G_t w^t, \quad (1)$$

or

$$x^{t+1} = \phi_t x^t + G_t w^t = (A_t + I)x^t + G_t w^t, \quad (2)$$

where x^t is the n -component state vector, w^t is an m -component system noise vector, A_t is an $n \times n$ matrix, and G_t is an $n \times m$ matrix. The initial state x_0 is assumed to be a Gaussian (normal) random variable with mean and variance given by

$$E[x^0] = \bar{x}^0 \quad \text{and} \quad E[(x^0 - \bar{x}^0)(x^0 - \bar{x}^0)^T] = M_0. \quad (3)$$

A DYNAMIC STOCHASTIC SYSTEM CONT.

Here w^t is a Gaussian purely random sequence (Joseph and Tou, 1961) with

$$E[w^t] = \bar{w}^t \quad \text{and} \quad E[(w^t - \bar{w}^t)(w^\tau - \bar{w}^\tau)^T] = Q_t \delta_{t\tau}, \quad (4)$$

where

$$\delta_{t\tau} = \begin{cases} 0 & \text{if } t \neq \tau, \\ 1 & \text{if } t = \tau. \end{cases} \quad (5)$$

Thus, Q_t represents the covariance matrix of the random vector w^t , and w^t and w^τ are independent random variables for $t \neq \tau$. We also assume that the sequence w^t is independent of the initial condition x^0 , i.e.,

$$E[(w^t - \bar{w}^t)(x^0 - \bar{x}^0)] = 0. \quad (6)$$

A DYNAMIC STOCHASTIC SYSTEM CONT.

The process of measurement of the state variables x^t yields a k -dimensional vector y^t which is related to x^t by the transformation

$$y^t = H_t x^t + v^t, \quad (7)$$

where H_t is the state-to-measurement transformation matrix of dimension $k \times n$, and v^t is a Gaussian purely random sequence of measurement noise vectors having the following properties:

$$E[v^t] = 0, \quad E[v^t \cdot (v^\tau)^T] = R_t \delta_{t\tau}, \quad (8)$$

$$E[(w^t - \bar{w}^t) \cdot (v^\tau)^T] = 0, \quad E[(x_0 - \bar{x}^0) \cdot (\bar{v}^t)^T] = 0. \quad (9)$$

In (8) the matrix R_t is the covariance matrix for the random variable v^t .

A DYNAMIC STOCHASTIC SYSTEM CONT.

Given a sequence of observations y^1, y^2, \dots, y^t up to time t , we would like to obtain the maximum likelihood estimate of the state x^t , or equivalently, to find the minimum weighted least squares estimate. In order to derive the estimate \hat{x}^t of x^t , we require the use of the Bayes theorem and an application of calculus to find the unconstrained minimum of the weighted least squares function. It yields the following recursive procedure for finding the estimate \hat{x}^t :

$$\begin{aligned}\hat{x}^t &= \bar{x}^t + (P_t H_t^T R_t^{-1})(y^t - H_t \bar{x}^t) \\ &= \bar{x}^t + K_t (y^t - H_t \bar{x}^t)\end{aligned}\tag{10}$$

with

$$K_t = P_t H_t^T R_t^{-1}.\tag{11}$$

The mean \bar{x}^t and the matrix P_t are calculated recursively by means of

$$\bar{x}^t = (A_{t-1} + I)\hat{x}^{t-1} + G_{t-1}\bar{w}^{t-1}, \quad \bar{x}^0 \text{ given}, \quad (12)$$

$$P_t = (M_t^{-1} + H_t R_t^{-1} H_t)^{-1}, \quad (13)$$

where

$$M_t = (A_{t-1} + I)P_{t-1}(A_{t-1} + I)^{-1} + G_{t-1}Q_{t-1}G_{t-1}, \quad M_0 \text{ given}. \quad (14)$$

The procedure in expressions (10)-(14) is known as the *Kalman filter* for linear discrete-time processes.

The interpretation of (10) is that the estimate \hat{x}^t is equal to the mean value \bar{x} plus a correction term which is proportional to the difference between the actual measurement y^t and the predicted measurement $H_t\bar{x}^t$. It should be noted that

$$P_t = E([x^t - \hat{x}^t)(x^t - \hat{x}^t)^T],$$

where P_t is the variance of $(x^t - \hat{x}^t)$. It therefore is a measure of the risk of $(x - \hat{x}^t)$. Thus, the proportionality matrix K_t can be interpreted as the ratio between the transformed risk $(P_t H_t)$ and the risk R_t in the measurement. Because of this property of K_t , it is called the *Kalman gain* in the engineering literature.

THE CONTINUOUS-TIME CASE: A WHITE NOISE PROCESS

Consider a Gaussian random process $w(t)$ with $E[w(t)] = \bar{w}(t)$ and the autocorrelation matrix

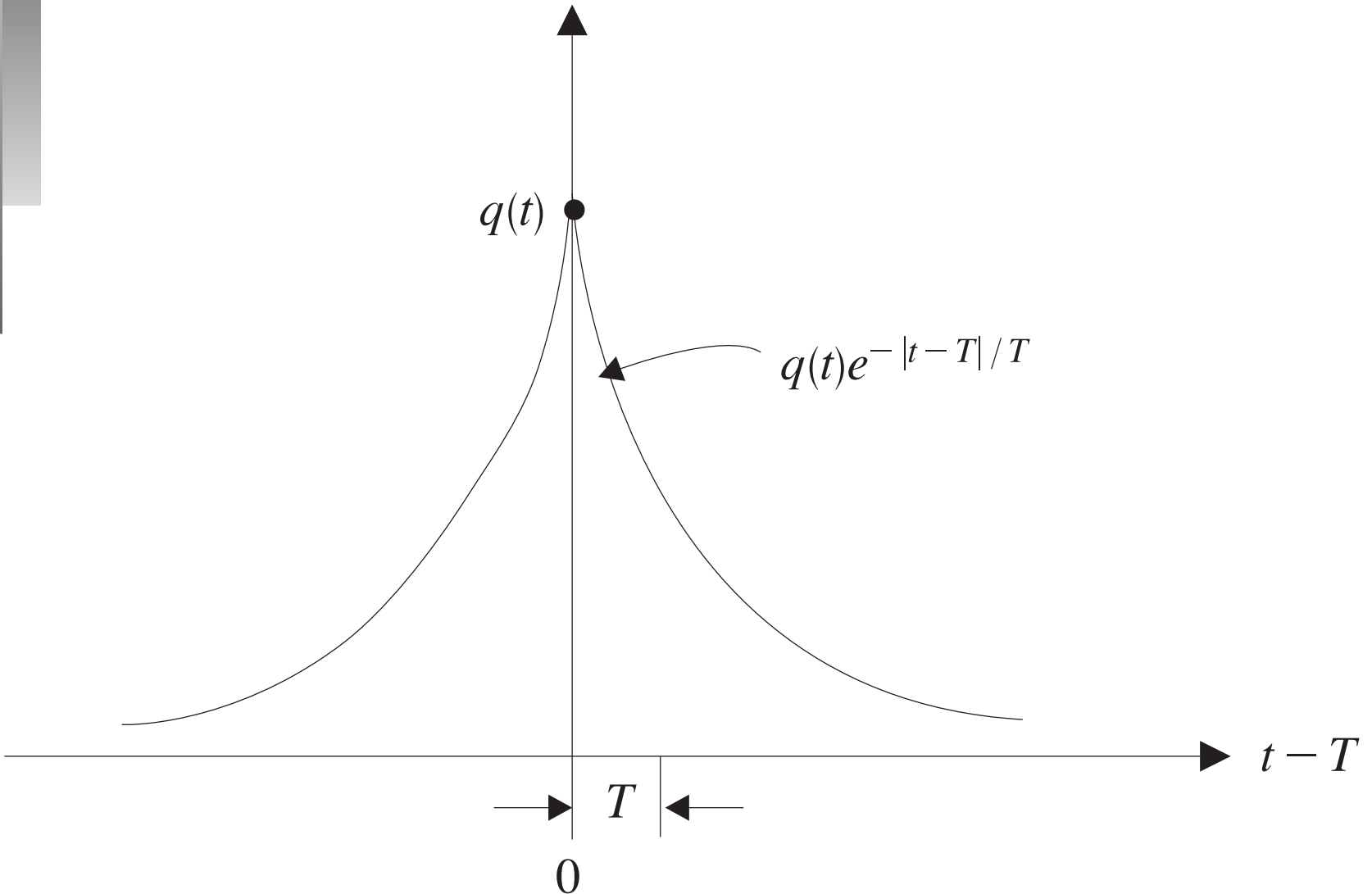
$$E[(w(t) - \bar{w}(t))(w(\tau) - \bar{w}(\tau))^T] = q(t)e^{-|t-\tau|/T}, \quad (15)$$

where the parameter T is small and $q(t)$ is the covariance matrix

$$q(t) = E[(w(t) - \bar{w}(t))(w(t) - \bar{w}(t))^T]. \quad (16)$$

Since T is small, it is obvious from (15) that the correlation between $w(t)$ and $w(\tau)$ decreases rapidly as $t - \tau$ increases. The autocorrelation function is sketched in Figure 13.1 for a scalar process.

FIGURE 13.1: AUTOCORRELATION FUNCTION FOR A SCALAR PROCESS



In order to obtain the white noise process, we let T approach 0, and define

$$Q(t) = \lim_{T \rightarrow 0, q(t) \rightarrow \infty} 2Tq(t). \quad (17)$$

Using the function $Q(t)$ we can approximate (15) by the expression

$$E[(w(t) - \bar{w}(t))(w(\tau) - \bar{w}(\tau))^T] = Q(t)\delta(t - \tau), \quad (18)$$

where $\delta(t - \tau)$, which is called the *Dirac delta function*, is the limit as $\varepsilon \rightarrow 0$ of

$$\delta_\varepsilon(t - \tau) = \begin{cases} 0 & \text{when } |t - \tau| > \varepsilon, \\ \frac{1}{2\varepsilon} & \text{when } |t - \tau| \leq \varepsilon. \end{cases} \quad (19)$$

The above discussion permits us to write formally the continuous analogue of (1) as

$$dx(t) = A(t)x(t)dt + G(t)w(t)dt. \quad (20)$$

Equation (20) can be restated as the linear *Itô stochastic differential equation*

$$dx(t) = A(t)x(t)dt + G(t)dz(t), \quad (21)$$

where $z(t)$ denotes a *Wiener process*; see Wiener (1949). Comparison of (20) and (21) suggests that $w(t)$ could be considered a “generalized derivative” of the Wiener process. A solution $x(t)$ of an Itô equation is a stochastic process, known as a *diffusion process*.

We also define the measurement process $y(t)$ in continuous time as

$$y(t) = H(t)x(t) + v(t), \quad (22)$$

where $v(t)$ is a white noise process with

$$E[v(t)] = 0 \quad \text{and} \quad E[v(t)v(\tau)^T] = R(t)\delta(t - \tau). \quad (23)$$

Using the theory of stochastic differential equations and calculus of maxima and minima, we can obtain a filter which provides the estimate $\hat{x}(t)$ given the observations $y(s)$, $s \in [0, t]$, as follows:

$$\dot{\hat{x}} = A\hat{x} + G\bar{w} + K[y - H\hat{x}], \quad \hat{x}(0) = E[x(0)], \quad (24)$$

$$K = PH^T R^{-1}, \quad (25)$$

$$\dot{P} = AP + PA^T - KHP + GQG^T, \quad (26)$$

$$P(0) = E[x(0)x^T(0)]. \quad (27)$$

This is called the *Kalman-Bucy filter* (Kalman and Bucy, 1961) for linear systems in continuous time.

13.2 STOCHASTIC OPTIMAL CONTROL

- For stochastic linear-quadratic optimal control problems (see Appendix D.1.1), 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 general stochastic optimal control problems, the separation principle does not hold. 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.

Let us consider the problem of maximizing

$$E\left[\int_0^T F(X_t, U_t, t)dt + S(X_T, T)\right], \quad (28)$$

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

$$dX_t = f(X_t, U_t, t)dt + G(X_t, U_t, t)dz_t, \quad X_0 = x_0. \quad (29)$$

ASSUMPTIONS AND NOTATION

For convenience in exposition we assume $F : E^1 \times E^1 \times E^1 \rightarrow E^1$, $S : E^1 \times E^1 \rightarrow E^1$, $f : E^1 \times E^1 \times E^1 \rightarrow E^1$ and $G : E^1 \times E^1 \times E^1 \rightarrow E^1$, so that (29) 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.

Since (29) 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 (29) is the same as $dz(t)$ in (21), except that in (29), dz_t is a scalar.

To solve the problem defined by (28) and (29), let $V(x, t)$, known as the *value function*, be the expected value of the objective function (28) from t to T , when an optimal policy is followed from t to T , given $X_t = x$. Then, by the principle of optimality,

$$V(x, t) = \max_u E[F(x, u, t)dt + V(x + dX_t, t + dt)]. \quad (30)$$

By Taylor's expansion, we have

$$\begin{aligned} V(x + dX_t, t + dt) = & V(x, t) + V_t dt + V_x dX_t + \frac{1}{2} V_{xx} (dX_t)^2 \\ & + \frac{1}{2} V_{tt} (dt)^2 + \frac{1}{2} V_{xt} dX_t dt + \text{higher-order terms.} \end{aligned} \quad (31)$$

From (29), we can formally write

$$(dX_t)^2 = f^2(dt)^2 + G^2(dz_t)^2 + 2fGdz_tdt, \quad (32)$$

$$dX_tdt = f(dt)^2 + Gdz_tdt. \quad (33)$$

For our purposes, it is sufficient to know the multiplication rules of the stochastic calculus:

$$(dz_t)^2 = dt, \quad dz_t \cdot dt = 0, \quad dt^2 = 0. \quad (34)$$

Substitute (31) into (30) and use (32), (33), (34), and the property that $E[z_t] = 0$ to obtain

$$V = \max_u E \left[Fdt + V + V_tdt + V_x fdt + \frac{1}{2} V_{xx} G^2 dt + o(dt) \right]. \quad (35)$$

Note that we have suppressed the arguments of the functions involved in (35).

Cancelling the the term V on both sides of (35), dividing the remainder by dt , and letting $dt \rightarrow 0$, we obtain the Hamilton-Jacobi-Bellman equation

$$0 = \max_u [F + V_t + V_x f + \frac{1}{2} V_{xx} G^2] \quad (36)$$

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

$$V(x, T) = S(x, T). \quad (37)$$

13.3 A STOCHASTIC PRODUCTION PLANNING MODEL

- X_t = the inventory level at time t (state variable),
 U_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,
 \bar{x} = the factory-optimal inventory level,
 \bar{u} = the factory-optimal production level,
 x_0 = the initial inventory level,
 h = the inventory holding cost coefficient,
 c = the production cost coefficient,
 B = the salvage value per unit of inventory at time T ,
 z_t = the standard Wiener process,
 σ = the constant diffusion coefficient.

The stock-flow equation is

$$dX_t = (U_t - S)dt + \sigma dz_t, \quad X_0 = x_0, \quad (38)$$

where x_0 denotes the initial inventory level. As in (20) and (21), 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

$$\min E \left\{ \int_0^T [c(U_t - \bar{u})^2 + h(X_t - \bar{x})^2] dt + BX_T \right\}. \quad (39)$$

Note that we do not restrict the production rate to be non-negative as required in Chapter 6. In other words, we permit disposal (i.e., $U_t < 0$). 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 OBJECTIVE FUNCTION CONT.

The solution of the above model will be carried out via the previous development of the Hamilton-Jacobi equation satisfied by a certain *value function*. To simplify the mathematics, we assume that

$$\bar{x} = \bar{u} = 0 \quad \text{and} \quad h = c = 1. \quad (40)$$

This assumption results in no loss of generality as the following analysis can be extended in a parallel manner for the case without (40). With (40), we restate (39) as

$$\max E \left\{ \int_0^T -(U_t^2 + X_t^2) dt + BX_T \right\}. \quad (41)$$

Let $V(x, t)$ denote the expected value of the objective function from time t to the horizon T with $X_t = x$ 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 Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \max_u [-(u^2 + x^2) + V_t + V_x(u - S) + \frac{1}{2}\sigma^2 V_{xx}] \quad (42)$$

with the boundary condition

$$V(x, T) = Bx. \quad (43)$$

Note that these are applications of (36) and (37) to the production planning problem.

It is now possible to maximize the expression inside the bracket of (28) with respect to u by taking its derivative with respect to u and setting it to zero. This procedure yields

$$u(x, t) = \frac{V_x(x, t)}{2}. \quad (44)$$

Substituting (44) into (42) yields the equation

$$0 = \frac{V_x^2}{4} - x^2 + V_t - SV_x + \frac{1}{2}\sigma^2 V_{xx}, \quad (45)$$

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 (43). The solution of (45) is considered in the next section.

It is important to remark that if production rate were restricted to be nonnegative, then (44) would be changed to

$$u(x, t) = \max \left[0, \frac{V_x(x, t)}{2} \right]. \quad (46)$$

Substituting (46) into (43) would give us a partial differential equation which must be solved numerically. We shall not consider (46) further in this chapter.

To solve equation (45) we let

$$V(x, t) = Q(t)x^2 + R(t)x + M(t). \quad (47)$$

Then,

$$V_t = \dot{Q}x^2 + \dot{R}x + \dot{M}, \quad (48)$$

$$V_x = 2Qx + R, \quad (49)$$

$$V_{xx} = 2Q, \quad (50)$$

where \dot{Y} denotes dY/dt . Substituting (48) in (45) and collecting terms gives

$$x^2[\dot{Q} + Q^2 - 1] + x[\dot{R} + RQ - 2SQ] + \dot{M} + \frac{R^2}{2} - RS + \sigma^2 Q = 0. \quad (51)$$

Since (51) must hold for any value of x , we must have

$$\dot{Q} = 1 - Q^2, \quad Q(T) = 0, \quad (52)$$

$$\dot{R} = 2SQ - RQ, \quad R(T) = B, \quad (53)$$

$$\dot{M} = RS - \frac{R^2}{4} - \sigma^2 Q, \quad M(T) = 0, \quad (54)$$

where the boundary conditions for the system of simultaneous differential equations (52), (53), and (54) are obtained by comparing (47) with the boundary condition $V(x, T) = Bx$ of (43).

To solve (52), 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

$$Q = \frac{y - 1}{y + 1}, \quad (55)$$

where

$$y = e^{2(t-T)}. \quad (56)$$

Since S is assumed to be a constant, we can reduce (53) 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

$$R = 2S + \frac{2(B - 2S)\sqrt{y}}{y + 1}. \quad (57)$$

Having obtained solutions for R and Q , we can easily express (54) as

$$M(t) = - \int_t^T [R(\tau)S - (R(\tau))^2/4 - \sigma^2 Q(\tau)] d\tau. \quad (58)$$

The optimal control is defined by (44), and the use of (55) and (57) yields

$$u^* = V_x/2 = Qx + R/2 = S + \frac{(y-1)x + (B-2S)\sqrt{y}}{y+1}. \quad (59)$$

REMARK 13.2

The optimal production rate in (59) 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.

If the demand rate S were time-dependent, it would have changed the solution of (53). Having computed this new solution in place of (57), we can once again obtain the optimal control as $u^* = Qx + R/2$.

Note that when $T \rightarrow \infty$, we have $y \rightarrow 0$ and

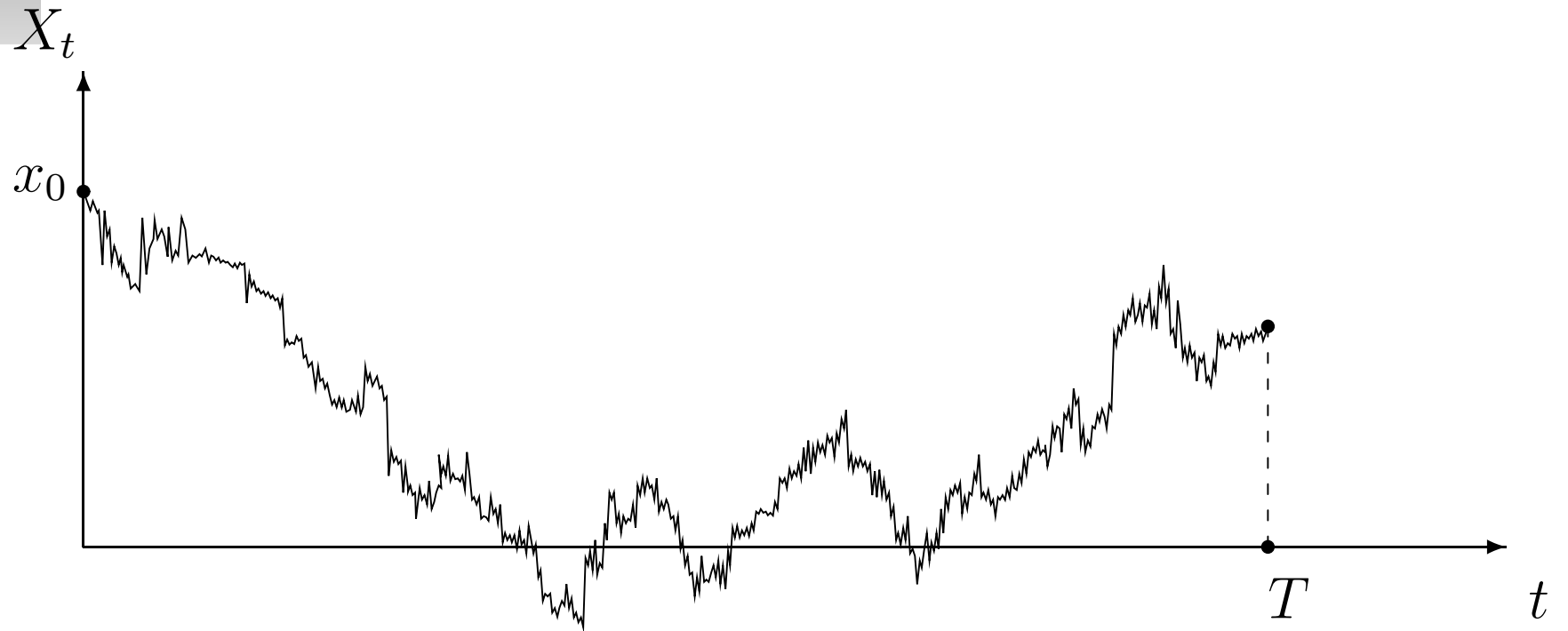
$$u^* \rightarrow S - x, \quad (60)$$

but the undiscounted objective function value (41) 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. This is carried out in Sethi and Thompson (1980).

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 13.2. 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 13.2: A SAMPLE PATH OF X_t
WITH $X_0 = x_0 > 0$ AND $B > 0$



13.4 A STOCHASTIC ADVERTISING PROBLEM

The stochastic advertising model due to Sethi (1983b) is :

$$\left\{ \begin{array}{l} \max E \left[\int_0^{\infty} e^{-\rho t} (\pi X_t - U_t^2) dt \right] \\ \text{subject to} \\ dX_t = (rU_t \sqrt{1 - X_t} - \delta X_t) dt + \sigma(X_t) dz_t, \quad X_0 = x_0, \\ U_t \geq 0, \end{array} \right. \quad (61)$$

where X_t is the market share and U_t is the rate of advertising at time t , and where the other parameters are as specified in Section 7.2.1. Note that 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 (61) is, therefore, to maximize the expected value of the total discounted profits.

An important consideration in choosing the function $\sigma(x)$ should be that the solution X_t to the Itô equation in (61) remains inside the interval $[0, 1]$. In addition to $x_0 \in (0, 1)$, it is assumed that

$$\sigma(x) > 0, \quad x \in (0, 1) \text{ and } \sigma(0) = \sigma(1) = 0. \quad (62)$$

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

$$u(x) \geq 0, \quad x \in (0, 1], \text{ and } u(0) > 0, \quad (63)$$

the Itô equation in (61) 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 (63), 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 13.2, we can write the HJB equation as

$$\rho V(x) = \max_u \left[\pi x - u^2 + V_x(ru\sqrt{1-x} - \delta x) + V_{xx}\sigma^2(x)/2 \right]. \quad (64)$$

Maximization of the RHS of (64) can be accomplished by taking its derivative with respect to u and setting it to zero.

This gives

$$u(x) = \frac{rV_x\sqrt{1-x}}{2}. \quad (65)$$

Substituting of (65) in (64) and simplifying the resulting expression yields the HJB equation

$$\rho V(x) = \pi x + \frac{V_x^2 r^2 (1-x)}{4} - V_x \delta x + \frac{1}{2} \sigma^2(x) V_{xx}. \quad (66)$$

SOLUTION OF THE HJB EQUATION

As shown in Sethi (1983b), a solution of (66) is

$$V(x) = \bar{\lambda}x + \frac{\bar{\lambda}^2 r^2}{4\rho}, \quad (67)$$

where

$$\bar{\lambda} = \frac{\sqrt{(\rho + \delta)^2 + r^2 \pi} - (\rho + \delta)}{r^2/2}, \quad (68)$$

as derived in Exercise 7.40. In Exercise 13.4, you are asked verify that (67) and (68) solve the HJB equation (66).

THE OPTIMAL FEEDBACK CONTROL

We can now obtain the optimal feedback control as

$$u^*(x) = \frac{r\bar{\lambda}\sqrt{1-x}}{2}. \quad (69)$$

Note that $u^*(x)$ satisfies the conditions in (63). It is easy to characterize (69) as

$$U_t^* = u^*(X_t) = \begin{cases} > \bar{u} & \text{if } X_t < \bar{x}, \\ = \bar{u} & \text{if } X_t = \bar{x}, \\ < \bar{u} & \text{if } X_t > \bar{x}, \end{cases} \quad (70)$$

where

$$\bar{x} = \frac{r^2\bar{\lambda}/2}{r^2\bar{\lambda}/2 + \delta} \quad (71)$$

and

$$\bar{u} = \frac{r\bar{\lambda}\sqrt{1-\bar{x}}}{2}. \quad (72)$$

THE OPTIMAL MARKET SHARE TRAJECTORY

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 (61). 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.

In Example 1.3, 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 $r > 0$.

We now 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 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 assume an infinite horizon problem in lieu of the bequest, for convenience in exposition.

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 shall rewrite as

$$dS_t = rS_t dt, \quad S_0 \text{ given.} \quad (73)$$

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

$$\frac{dP_t}{P_t} = \alpha dt + \sigma dz_t, \quad P_0 \text{ given}, \quad (74)$$

or simply,

$$dP_t = \alpha P_t dt + \sigma P_t dz_t, \quad P_0 \text{ given}, \quad (75)$$

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.

The LHS in (74) 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 (74) is often referred to as a *logarithmic Brownian Motion*.

In order to complete the formulation, we need the following additional notation:

W_t = the wealth at time t ,

C_t = the consumption rate at time t ,

Q_t = the fraction of the wealth invested in stock at time t ,

$1 - Q_t$ = the fraction of the wealth kept in the savings account at time t ,

$U(c)$ = the utility of consumption when consumption is at the rate c ; the function $U(c)$ is assumed to be increasing and concave,

ρ = the rate of discount applied to consumption utility,

B = the bankruptcy parameter to be explained later.

We write the wealth equation informally as

$$\begin{aligned}
 dW_t &= Q_t W_t \alpha dt + Q_t W_t \sigma dz_t + (1 - Q_t) r W_t dt - C_t dt & (76) \\
 &= (\alpha - r) Q_t W_t dt + (r W_t - C_t) dt + \sigma Q_t W_t dz_t, \quad W_0 \text{ given.}
 \end{aligned}$$

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 $(1 - Q_t) r W_t dt$ 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 (76), 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 (76) from (73) and (74), 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 shall 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:

$$\max \left\{ J = E \left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\}, \quad (77)$$

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 .

THE OPTIMAL CONTROL PROBLEM

Let us recapitulate the optimal control problem of Rich Investor:

$$\left\{ \begin{array}{l} \max \left\{ J = E \left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\} \\ \text{subject to} \\ dW_t = (\alpha - r)Q_t W_t dt + (rW_t - C_t)dt + \sigma Q_t W_t dt, \\ C_t \geq 0. \end{array} \right. \quad \begin{array}{l} (78) \\ W_0 \text{ given,} \end{array}$$

As in the infinite horizon problem of Section 13.3, 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 13.2, the HJB equation satisfied by the value function $V(x)$ for problem (78) can be written as

$$\begin{cases} \rho V(x) = \max_{c \geq 0, q} [(\alpha - r)qxV_x + (rx - c)V_x \\ \quad + (1/2)q^2\sigma^2x^2V_{xx} + U(c)], \\ V(0) = B. \end{cases} \quad (79)$$

ADDITIONAL ASSUMPTIONS

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

$$U(c) = \ln c. \quad (80)$$

This utility has an important simplifying property, namely,

$$U'(0) = 1/c|_{c=0} = \infty. \quad (81)$$

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

ADDITIONAL ASSUMPTIONS CONT.

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 (13.81). While we have provided an intuitive explanation, it is possible to show rigorously that condition (81) together with $B = -\infty$ implies a strictly positive consumption level at all times and no bankruptcy.

SOLUTION OF THE HBJ EQUATION

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 (79) 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.,

$$q(x) = -\frac{(\alpha - r)V_x}{x\sigma^2V_{xx}}, \quad (82)$$

and

$$c(x) = \frac{1}{V_x}. \quad (83)$$

SOLUTION OF THE HBJ EQUATION CONT.

Substituting (82) and (83) in (79) allows us to remove the max operator from (79), and provides us with the equation

$$\rho V(x) = -\frac{\gamma(V_x)^2}{V_{xx}} + (rx - 1/V_x)V_x - \ln V_x, \quad (84)$$

where

$$\gamma = \frac{\alpha - r}{2\sigma^2}. \quad (85)$$

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 (84) into a second-order, linear, ordinary differential equation.

SOLUTION OF THE HBJ EQUATION CONT.

Assume that the value function is strictly concave and, therefore, V_x is monotonically decreasing in x . This means that the function $c(\cdot)$ defined in (83) has an inverse $X(\cdot)$ such that (84) can be rewritten as

$$\rho V(X(c)) = -\frac{\gamma(U'(c))^2 X'(c)}{U''(c)} + (rX(c) - c)U'(c) + U(c). \quad (86)$$

Differentiation with respect to c yields the intended second-order, linear ordinary differential equation

$$\begin{aligned} \gamma X''(c) &= \left[(r - \rho - 2\gamma) \frac{U''(c)}{U'(c)} + \frac{\gamma U'''(c)}{U''(c)} \right] X'(c) \\ &+ \left[\frac{U''(c)}{U'(c)} \right]^2 (rX(c) - c). \end{aligned} \quad (87)$$

SOLUTION OF THE HBJ EQUATION CONT.

This equation has an explicit solution with three parameters to be determined; see Appendix A. After some calculations, one can determine these parameters, and obtain the solution of (84) as

$$V(x) = \frac{1}{\rho} \ln(\rho x) + \frac{r - \rho + \gamma}{\rho^2}, \quad x \geq 0. \quad (88)$$

In Exercise 13.3, you are asked by a direct substitution in (84) to verify that (88) is indeed a solution of (84).

Moreover, $V(x)$ defined in (88) is strictly concave, so that our concavity assumption made earlier is justified.

THE OPTIMAL FEEDBACK CONTROL

From (88), it is easy to show that (82) and (83) yield the following feedback policies:

$$q^*(x) = \frac{\alpha - r}{\sigma^2}, \quad (89)$$

$$c^*(x) = \rho x. \quad (90)$$

The investment policy (89) says that the optimal fraction of the wealth invested in the risky stock is $(\alpha - r)/\sigma^2$, i.e.,

$$Q_t^* = q^*(W_t) = \frac{\alpha - r}{\sigma^2}, \quad t \geq 0, \quad (91)$$

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

$$C_t^* = c^*(W_t) = \rho W_t, \quad t \geq 0. \quad (92)$$

13.6 CONCLUDING REMARKS

- Impulse stochastic control:
Bensoussan and Lions (1984).
- Stochastic control problems with jump Markov processes or martingale problems:
Fleming and Soner (1992), Davis (1993), and Karatzas and Shreve (1998).
- Applications to manufacturing problems:
Sethi and Zhang (1994a) and Yin and Zhang (1997).

CONCLUDING REMARKS CONT.

- Applications to finance:
Sethi (1997a) and Karatzas and Shreve (1998).
- Applications to marketing:
Tapiero (1988), Raman (1990), and Sethi and Zhang (1995).
- Applications to economics:
Pindyck (1978a, 1978b), Rauser 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).