CHAPTER 12

DYNAMIC PROGRAMMING

In science, what is susceptible to proof must not be believed without proof.

— *R. Dedekind (1887)*

This chapter gives an introduction to the theory of dynamic optimization. The term *dynamic* is used because the problems involve systems that evolve over time. Time is measured by the number of whole periods (say weeks, quarters, or years) that have passed since time 0, so we speak of *discrete time*. In this case, it is natural to study dynamical systems whose development is governed by difference equations.

If the horizon is finite, dynamic optimization problems can be solved, in principle, using classical calculus methods. There are, however, solution techniques that take advantage of the special structure of discrete dynamic optimization problems. Section 12.1 introduces a method that converts any, potentially very complicated, dynamic problem into an equivalent series of much simpler problems. This *nesting* technique was given the name *dynamic programming* by American mathematician Richard Bellman, who is credited with its development. Section 12.2 gives a dynamic programming version of the Euler equation used in continuous time control theory, which is much used in the economics literature.

Section 12.3 presents the fundamental result in dynamic programming with an infinite time horizon, the Bellman equation. Section 12.4 offers an introduction to stochastic dynamic programming, including the stochastic Euler equation, while the concluding section is devoted to the important case of stationary problems with an infinite horizon.

When a discrete time dynamic optimization problem has restrictions on the terminal values of the state variable, there is a discrete time version of the maximum principle which may work better than dynamic programming. An appendix to this chapter sets out the relevant discrete time maximum principle. This appendix also presents a brief discussion of infinite horizon problems in this setting.

12.1 Dynamic Programming in Finite Horizon

Consider a system that changes at discrete times $t = 0, 1, \ldots, T$. Suppose that the **state** of the system at time t is characterized by a real number x_t.[1] Assume that the initial state x_0 is historically given, and that from then on the system evolves through time under the influence of a sequence of **controls** u_t.[2] The controls influence the evolution of the system through a difference equation

$$x_{t+1} = g_t(x_t, u_t),\ x_0 \text{ given},\ \ u_t \in U_t(x_t),\ t = 0, \ldots, T-1. \tag{1}$$

where, for each t, g_t is a *given* function. Thus, we assume that the state of the system at time $t+1$ depends on the time t, on the state x_t in the preceding period t, and on u_t, the value chosen for the control at time t. Assume also that the set of controls that is feasible at each period depends on the specific period and on the state. That is, suppose that at time t, u_t can be chosen freely from set $U_t(x_t)$, which is called the **control region**

Suppose that we choose values for $u_0, u_1, \ldots, u_{T-1}$. Then, Equation (1) immediately gives $x_1 = g_0(x_0, u_0)$. Since x_1 is now determined, so too is $x_2 = g_1(x_1, u_1)$, then $x_3 = g_2(x_2, u_2)$, and so on. In this way, (1) can be used to compute recursively the successive states $x_1, x_2, \ldots, x_T$ in terms of the initial state, x_0, and the controls. In this way, each choice of controls $(u_0, u_1, \ldots, u_{T-1})$ gives rise to a sequence of states $(x_1, x_2, \ldots, x_T)$. Let us denote corresponding pairs $(x_0, \ldots, x_T)$, $(u_0, \ldots, u_T)$ by $(\{x_t\}, \{u_t\})$, and call them **admissible sequence pairs**.

Different paths will usually have different utility or value for a given decision maker: the problem assumes that in each time period, she enjoys a utility level that depends on the state and the control in that period. The whole path is then assessed according to the aggregate of these per-period utilities.[3] Formally, assume that for each t there is a **return function** $f_t(x, u)$ such that the utility associated with a given path is represented by the sum $\sum_{t=0}^{T} f_t(x_t, u_t)$. This sum is called the **objective function**.[4]

For each admissible sequence pair the objective function then has a definite value. We shall study the following problem: *Among all admissible sequence pairs* $(\{x_t\}, \{u_t\})$, *find one, denoted by* $(\{x_t^*\}, \{u_t^*\})$, *that makes the value of the objective function as large as possible.*

The admissible sequence pair $(\{x_t^*\}, \{u_t^*\})$ is called an **optimal pair**, and the corresponding control sequence $\{u_t^*\}_{t=0}^{T}$ is called an **optimal control**.

The discrete time optimization problem can be formulated briefly as

$$\max \sum_{t=0}^{T} f_t(x_t, u_t) \text{ subject to } x_{t+1} = g_t(x_t, u_t),\ u_t \in U_t(x_t);\ x_0 \text{ given}. \tag{2}$$

---

[1] For example, x_t might be the quantity of grain that is stockpiled in a warehouse at time t.

[2] In the context of the example above, u_t would be the quantity of grain removed from the stock at time t.

[3] Again, in the context of the example, $f_t(x, u)$ would be the net profit that the agent obtains from selling quantity u, removed from stock x.

[4] The return function is often called *performance*, *felicity* or *instantaneous utility* function. The objective function is often referred to as *intertemporal utility* function. The objective function is sometimes specified as $\sum_{t=0}^{T-1} f_t(x_t, u_t) + S(x_T)$, where S measures the net value associated with the terminal period. This is a special case in which $f_T(x_T, u_T) = S(x_T)$, and S is often called a *scrap value function.*

EXAMPLE 1: Let x_t be an individual's wealth at time t. At each time period t, the individual has to decide what proportion, u_t, of his wealth to consume, leaving the remaining proportion for savings. Assume that wealth earns interest at the constant rate $\rho - 1 > 0$ each period, so that her wealth is $x_{t+1} = \rho(1 - u_t)x_t$ at the beginning of period $t + 1$. This equation holds for $t = 0, \ldots, T - 1$, with $x_0 > 0$ given. Suppose that the utility of consuming c at time t is $W_t(c)$. Then, the total utility is $\sum_{t=0}^{T} W_t(u_t x_t)$, and the problem facing the individual is

$$\max \sum_{t=0}^{T} W_t(u_t x_t) \text{ subject to } x_{t+1} = \rho(1 - u_t)x_t,\ u_t \in [0, 1];\ x_0 \text{ given.} \qquad (*)$$

(See Problems 2, 3, and 8.) □

12.1.1 An Illustrative Problem.

Before giving the formal analysis of the technique developed by Bellman, it is useful to present a simple application that will help us develop the intuition that lies behind the formal analysis. Consider Figure 1.a, and suppose that each of the nodes in that graph represents a city, and that the arrows represent one-way roads that connect these cities. Suppose that a tractor salesman based in source city σ has to go on a sales trip that ends at terminal city τ. Being a good salesperson, he will take this opportunity to sell his tractors to farmers on the roads that he takes on the trip. The numbers that are next to each arrow in the graph denote the number of tractors that our salesman would sell in that part of his trip. His optimization problem is, then, to find the route that takes him from σ to τ and maximizes the number of tractors he will sell on the trip.[5]

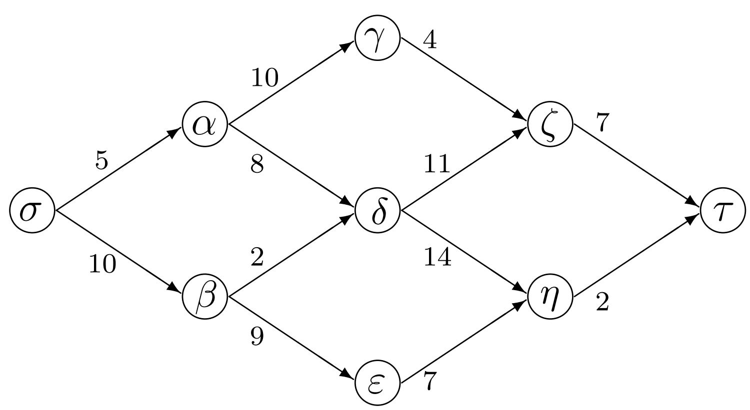

Figure 1.a

Note that when he is at source σ, if he only looks at the tractors to be sold in the first leg of the trip, he will chose to go to city β first. Of course, this *short-sighted* decision need not be optimal: by choosing to go to β, he is ruling out the possibility of ever going to city γ, and of selling tractors to farms that are in the roads that connect α with γ and with δ. The salesman has to be more thoughtful. One possibility would be for him to consider all

[5]Remember that these roads are all one way, so he cannot loop around and cover all farms.

possible combinations of cities and the resulting sales. But is there a better way to find the optimal solution?

It turns out that our salesman can find his optimal route in a simpler way, requiring less computations and obtaining more information than if he solved his problem by finding all possible combinations. Instead of solving only the problem of going from σ to τ, all he has to do, following Bellman's intuition, is to solve a series of simpler problems, looking at the best way to go from *any* city to τ, starting from those that are closest to τ.

Suppose that he finds himself at ζ. There are no options for him, and he will sell seven tractors in the last leg of his trip. If he finds himself at η, there is nothing to decide and he will sell two. These two *subproblems* were very easy to solve. If he finds himself at cities γ or ε in the penultimate leg, his problems is equally simple, since he has no choices: from γ he'll have to go via ζ, selling 11 tractors in total; from ε, he'll sell nine tractors going via η. If he finds himself at δ, on the other hand, he has a decision to make—whether to go via ζ or via η. His problem is slightly more complicated, for he has to consider the tractors he will sell in the two remaining legs of his trip. Looking at that subproblem, it is evident that the short-sighted decision to go via η would be a mistake: considering the subsequent leg, he should go via ζ, selling 18 tractors in total.

Note that what we have done so far gives us the optimal routes the salesman should take from any city in $\{\delta, \ldots, \eta\}$ to τ. Now, suppose that he finds himself at α. He does have a decision to make, and this decision will imply not only a number of sales made in the immediate leg, but also will restrict the number of sales he will *optimally* make in the rest of his trip. His problem now looks more complicated. But since he has already solved the subproblems for all subsequent legs, he can be clever and reason his way to a simpler problem. In particular, he knows that if he goes from α to δ, he will not choose to go to η afterwards, so he can dismiss that route and make his problem simpler. Effectively, he can use his previous reasoning of the optimal subsequent choices, to delete the pre-terminal cities from his map, as in Figure 1.b.

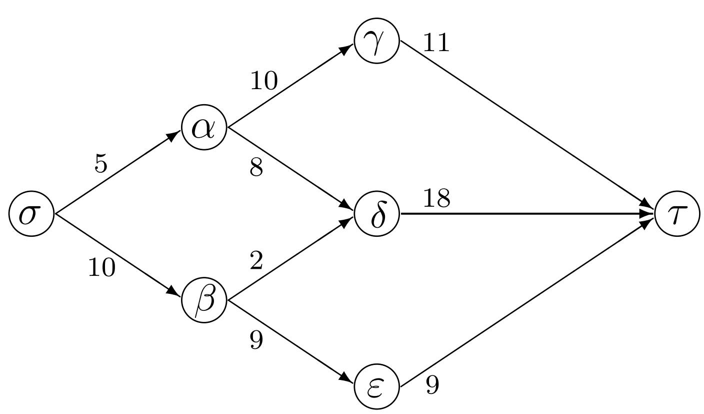

Figure 1.b

Looked at it in this new graph, the decision problem at α is much easier, and it is obvious that he should go via δ, selling 26 tractors. If he were to start from β, it is also immediate from Figure 1.b that the salesman would find it optimal to sell 20 tractors, going via δ.

Now, let us consider the problem starting from σ. We have already solved the problems starting from α and from β, and know that the optimal sale levels are 26 and 20 tractors respectively. The sales person thus knows that he will *not* go to cities γ and ε, so that we can delete these cities from the map and simply consider Figure 1.c.

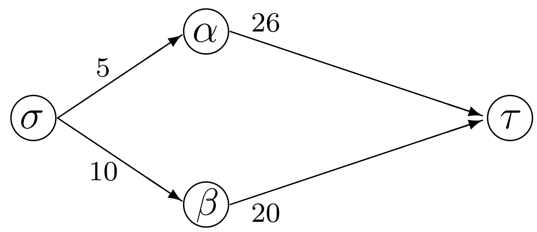

Figure 1.c

It is now clear that choosing to first go to β would have been a mistake, for the maximum number of sales he would have been able to make afterwards would have limited his total sales to 30 tractors. The best choice for the salesman is to go via α, thus making a total of 31 sales on his way to τ.

Let us now cast this situation in the language of the general problem. There are four legs to the salesman's trip. For $t = 1, \dots, 4$, let x_t represent the city where the he is after the t-th leg. This city will be the state of the problem after that leg, and the initial state is the source city, $x_0 = \sigma$. At the beginning of the problem, he can choose to go 'up' or 'down'. These are the values of his control variable. If starting at σ he goes up, he will start the next leg at α. In the general language, $U_1(\sigma) = \{\text{up}, \text{down}\}$, $g_0(\sigma, \text{up}) = \alpha$, and $g_0(\sigma, \text{down}) = \beta$. The sets of controls and the dynamics of the states are constructed similarly from other cities, with care that $U_2(\gamma) = U_3(\zeta) = \{\text{down}\}$, and $U_2(\varepsilon) = U_3(\eta) = \{\text{up}\}$,

The return function will be the tractors to be sold in each leg, as a function of where the salesman begins that leg and what choice he makes. For instance, $f_0(\text{up}, \sigma) = 5$ and $f_1(\text{down}, \alpha) = 8$, as in Figure 1.d. The decision facing him is Problem (2).[6]

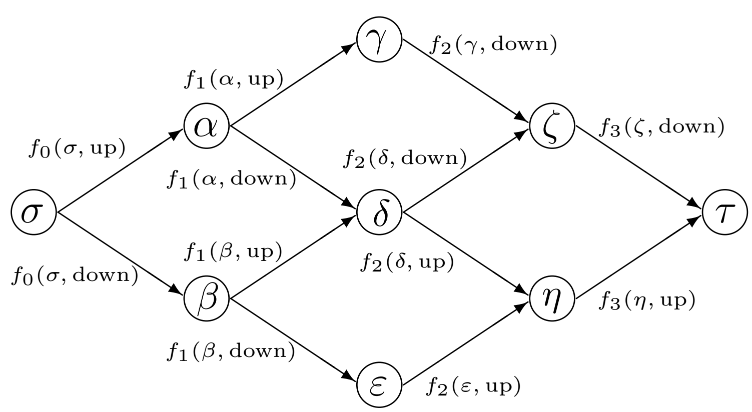

Figure 1.d

[6]With the proviso that $f_4(\tau) = 0$.

Our next step will be to present a general method for the solution of (2); we will then point out how what we did in this section is just an application of that method.

12.1.2 The Value Function and its Properties.

Suppose that at time $s < T$ the state of the system is x. The best that can be done in the remaining periods is to choose $u_s, u_{s+1}, \ldots, u_T$, and thereby also $x_{s+1}, x_{s+2}, \ldots, x_T$, to maximize $\sum_{t=s}^{T} f_t(x_t, u_t)$, with $x_s = x$. We define the **(optimal) value function** for the problem at time $s < T$ by

$$J_s(x) = \max \sum_{t=s}^{T} f_t(x_t, u_t) \text{ subject to } x_{t+1} = g_t(x_t, u_t),\ u_t \in U_t(x_t);\ x_s = x, \tag{3}$$

where we assume that the maximum is attained.[7]

At time T, if the state of the system is x, the best that can be done is simply to maximize $f_T(u, x)$ by our choice of u. If such optimum exists, we define the value function at T as

$$J_T(x) = \max f_T(x, u), \text{ subject to } u \in U_T(x). \tag{4}$$

Two features in this construction are worth emphasizing. If $U_t(x)$ is empty, then, we will adopt the convention that the maximum over $U_t(x)$ is set equal to $-\infty$.[8] Now, let $X_t(x_0)$ denote the range of all possible values of the state x_t that can be generated by the difference equation (1), if we start in state x_0 and then go through all possible values of $u_0, u_1, \ldots, u_{t-1}$. Of course, only the values of $J_t(x)$ for $x \in X_t(x_0)$ are relevant for the domain of the value functions.

We now prove an important property of the value function. Suppose that at a particular time $s < T$ we find ourselves in state $x_s = x$. What is the optimal choice for u_s? If we choose $u_s = u$, then at time σ we obtain the immediate reward $f_s(x, u)$, and, according to (3), the state at time $s+1$ will be $x_{s+1} = g_s(x, u)$. Using definition (3) again, the highest attainable value of the total reward from time $s+1$ to time T, $\sum_{t=s+1}^{T} f_t(x_t, u_t)$ starting from the state x_{s+1}, is $J_{s+1}(x_{s+1}) = J_{s+1}[g_s(x, u)]$. Hence, the best choice of the control at time s must be a value of u that maximizes the sum

$$f_s(x, u) + J_{s+1}[g_s(x, u)].$$

This leads to the following general result:

THEOREM 1 (Fundamental Equations of Dynamic Programming). *For each $s = 0, 1, \ldots, T-1$, let $J_s(x)$ be the value functions (3) for Problem (2). The sequence of value functions satisfies the equations*

$$J_s(x) = \max_{u \in U_s(x)} \{f_s(x, u) + J_{s+1}[g_s(x, u)]\}, \tag{5}$$

---

[7] This is true if, for example, the functions f and g are continuous and U is compact.

[8] Frequently, the set $U_t(x)$ is determined by one or more inequalities of the form $h_t(x, u) \leq 0$, for some continuous function h_t.

for $s = 0, 1, \ldots, T-1$.

Theorem 1 gives us the basic tool for solving dynamic optimization problems. It is used as follows: First, find the function $J_T(x)$ by using its definition, (5). The maximizing value of u depends (usually) on x, and is denoted by $u_T^*(x)$. The next step is to use (4) to determine $J_{T-1}(x)$ and the corresponding maximizing control $u_{T-1}^*(x)$. Then, work backwards in this manner to determine, recursively, all the value functions $J_{T-2}(x), \ldots, J_0(x)$ and the maximizers $u_{T-2}^*(x), \ldots, u_0^*(x)$.

This allows us to construct the solution of the original optimization problem: Since the state at $t = 0$ is x_0, a best choice of u_0 is $u_0^*(x_0)$. After $u_0^*(x_0)$ is found, the difference equation in (1) determines the state at time 1 as $x_1^* = g_0[x_0, u_0^*(x_0)]$. Then, $u_1^*(x_1^*)$ is a best choice of u_1, and this choice determines x_2^* by (1). Again, $u_2^*(x_2^*)$ is a best choice of u_2, and so on.[9]

EXAMPLE 2: We can now use the example presented in Subsection 12.1.1 to illustrate the recursive method of Theorem 1. Figure 1.d expressed the salesman problem in our general notation. Our first step was to realize that $J_3(\zeta) = f_3(\zeta, \text{down})$, and $J_3(\eta) = f_3(\eta, \text{up})$, as there were not choices available at that stage. Figure 2.a makes this explicit.

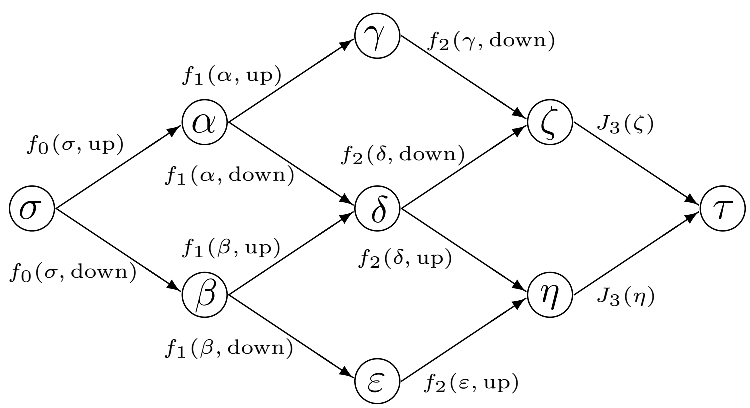

Figure 2.a

It is easy to see, from Figure 2.a, that our equally simple solution to the problems facing the salesman at cities γ and ε gave us that $J_2(\gamma) = f_2(\gamma, \text{down}) + J_3(\zeta)$, and $J_2(\varepsilon) = f_2(\varepsilon, \text{up}) + J_3(\eta)$. The interesting case was if he found himself at δ, for there he has a decision to make. Recalling our analysis there, it is clear that

$$J_2(\delta) = \max\{f_2(\delta, \text{up}) + J_3(\zeta), f_2(\delta, \text{down}) + J_3(\eta)\}.$$

Then, we can delete the penultimate stage of the problem, as in Figure 2.b.

[9] If we minimize rather than maximize the sum in (2), then Theorem 1 holds with "max" replaced by "min" in (3), (4) and (5). This is because minimizing f is equivalent to maximizing $-f$.

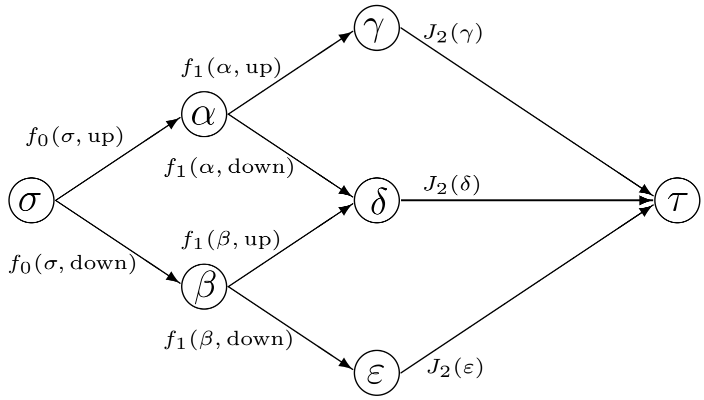

Figure 2.b

Again, we now have that

$$J_1(\alpha) = \max\{f_1(\alpha, \text{up}) + J_2(\gamma), f_1(\alpha, \text{down}) + J_2(\delta)\},$$

and

$$J_1(\beta) = \max\{f_1(\beta, \text{up}) + J_2(\delta), f_1(\beta, \text{down}) + J_2(\varepsilon)\},$$

which we can now introduce to Figure 2.c.

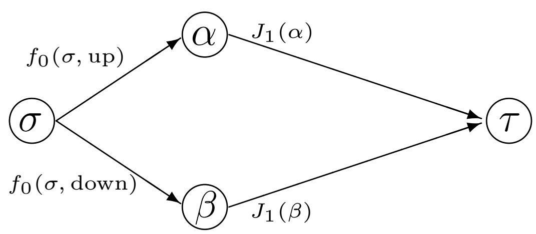

Figure 2.c

Finally, we conclude that the solution to Problem (2) for the salesman is simply

$$J_0(\sigma) = \max\{f_0(\sigma, \text{up}) + J_1(\alpha), f_0(\sigma, \text{down}) + J_1(\beta)\}.$$

EXAMPLE 3: Use Theorem 1 to solve the problem

$$\max \sum_{t=0}^{3} (1 + x_t - u_t^2) \text{ subject to } x_{t+1} = x_t + u_t,\ u_t \in \mathbb{R};\ x_0 = 0.$$

Solution: Here, the horizon is $T = 3$, while at all t the return function is $f_t(x, u) = 1 + x - u^2$ and the updating function is $g_t(x, u) = x + u$. Consider first Equation (5) and note that $J_3(x)$ is the maximum value of $1 + x - u^2$ for $u \in (-\infty, \infty)$. This maximum value is obviously attained for $u = 0$. Hence, using the notation introduced in Theorem 1,

$$J_3(x) = 1 + x, \text{ with } u_3^*(x) \equiv 0. \tag{$*$}$$

For $s = 2$, the function to be maximized in (4) is $1 + x - u^2 + J_3(x + u)$. Let us denote this function by $h_2(u)$, and use the fact that $(*)$ implies that $J_3(x + u) = 1 + (x + u)$, to get

$$h_2(u) = 1 + x - u^2 + 1 + (x + u) = 2 + 2x + u - u^2.$$

Function h_2 is concave in u, and $h_2'(u) = 1 - 2u$, which vanishes for $u = 1/2$, so this is the optimal choice of u. The maximum value of $h_2(u)$ is, thus,

$$h_2(\tfrac{1}{2}) = 2 + 2x + \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{9}{4} + 2x.$$

Hence,

$$J_2(x) = \tfrac{9}{4} + 2x, \quad \text{with } \; u_2^*(x) \equiv \tfrac{1}{2} \tag{$**$}$$

For $s = 1$, the function to be maximized in (7) is given by $h_1(u) = 1 + x - u^2 + J_2(x + u)$. Substituting from $(**)$ gives

$$h_1(u) = 1 + x - u^2 + \tfrac{9}{4} + 2(x + u) = \tfrac{13}{4} + 3x + 2u - u^2.$$

Because h_1 is concave and $h_1'(u) = 2 - 2u$, which is null for $u = 1$, the maximum value of $h_1(u)$ is $13/4 + 3x + 2 - 1 = 17/4 + 3x$. Hence,

$$J_1(x) = \tfrac{17}{4} + 3x, \quad \text{with } \; u_1^*(x) \equiv 1 \tag{$***$}$$

Finally, for $s = 0$, the function to be maximized is $h_0(u) = 1 + x - u^2 + J_1(x + u)$. Substituting from $(***)$ gives

$$h_0(u) = 1 + x - u^2 + \tfrac{17}{4} + 3(x + u) = \tfrac{21}{4} + 4x + 3u - u^2.$$

Function h_0 is concave and $h_0'(u) = 3 - 2u$ vanishes for $u = 3/2$, so the maximum value of $h_0(u)$ is $h_0(3/2) = 21/4 + 4x + 9/2 - 9/4 = 15/2 + 4x$. Thus,

$$J_0(x) = \tfrac{15}{2} + 4x, \quad \text{with } \; u_0^*(x) \equiv \tfrac{3}{2}$$

In this particular case, the optimal choice of the control each period is a constant, independent of the state. The corresponding optimal values of the state variables are $x_1 = x_0 + u_0 = 3/2$, $x_2 = x_1 + u_1 = 3/2 + 1 = 5/2$, $x_3 = x_2 + u_2 = 5/2 + 1/2 = 3$. The maximum value of the objective function is $15/2$. □

In principle, all deterministic finite horizon dynamic problems can be solved in an alternative way using ordinary calculus. But the method becomes very unwieldy if the horizon T is large, and is usually hopelessly impractical for stochastic optimization problems of the kind considered in Sections 12.4–12.5.

In the next example the terminal time is an arbitrary natural number, and the optimal control turns out to be a specific function of the state of the system.

EXAMPLE 4: Solve the following problem:

$$\max \left(\sum_{t=0}^{T-1} -\tfrac{2}{3} u_t x_t + \ln x_T \right) \text{ subject to } x_{t+1} = x_t(1 + u_t x_t),\ u_t \geq 0;\ x_0 > 0. \tag{7}$$

Solution: Because $x_0 > 0$ and $u_t \geq 0$, we have $x_t > 0$ for all t. Now, $f_T(x, u) = \ln x$ is independent of u, so $J_T(x) = \ln x$, and any u_T is optimal.

Next, putting $s = T - 1$ in (4) yields

$$J_{T-1}(x) = \max_{u \geq 0} \left\{ -\tfrac{2}{3} ux + J_T[x(1 + ux)] \right\} = \max_{u \geq 0} \left[-\tfrac{2}{3} ux + \ln x + \ln(1 + ux) \right].$$

The maximum of the concave function $h(u) = -\frac{2}{3} ux + \ln x + \ln(1 + ux)$ is at the point where its derivative is 0. This gives $h'(u) = -\frac{2}{3} x + x/(1 + ux) = 0$, or, since $x > 0$, $u = 1/2x$. Then $h(1/2x) = \ln x - 1/3 + \ln(3/2)$. Hence, $J_{T-1}(x) = h(1/2x) = \ln x + C$, with $C = -1/3 + \ln(3/2)$, and $u^*_{T-1}(x) = 1/2x$.

The next step is to use (4) for $s = T - 2$:

$$J_{T-2}(x) = \max_{u \geq 0} \left\{ -\tfrac{2}{3} ux + J_{T-1}[x(1 + ux)] \right\} = \max_{u \geq 0} \left[-\tfrac{2}{3} ux + \ln x + \ln(1 + ux) + C \right]$$

Again, $u = u^*_{T-2}(x) = 1/2x$ gives the maximum because the first-order condition is the same, and we get $J_{T-2}(x) = \ln x + 2C$, with $C = -1/3 + \ln(3/2)$, and $u^*_{T-2}(x) = 1/2x$. This pattern continues and so, for $k = 0, 1, \ldots, T$, we get $J_{T-k}(x) = \ln x + kC$, with $C = -1/3 + \ln(3/2)$ and $u^*_{T-k}(x) = 1/2x$.

So far we have been working backwards from time T to time 0. Putting $t = T - k$ for each k, we find that $J_t(x) = \ln x + (T - t)C$ and $u^*_2 = 1/2x$ for $t = 0, 1, \ldots, T$. Finally, inserting $u^*_t = 1/2x^*_t$ in the difference equation gives $x^*_{t+1} = (\frac{3}{2})x^*_t$. So $x^*_t = (\frac{3}{2})^t x_0$, with $\bar{u}_t = (\frac{2}{3})^t / 2x_0$ as optimal control values. □

In the above formulation, the state x and the control u may well be vectors, in say $\mathbb{R}^n$ and $\mathbb{R}^r$, respectively. Then g must be a vector function as well, and the difference equation is a system of difference equations, one for each component of x. No changes are then needed in Theorem 1, except that we would use boldface letters for x, u, and g.

Controls $u_t(x)$ that depend on the state x of the system are called **closed-loop controls**, whereas controls u_t that only depend on time are called **open-loop controls**.[10] Given the initial state x_0 and a sequence of closed-loop controls $u^*_t(x)$, the evolution of the state x_t is uniquely determined by the difference equation

$$x_{t+1} = g_t(x_t, u_t(x_t)),\ x_0 \text{ given.} \tag{$*$}$$

[10]Except in rare special cases, the controls $u^*_s, u^*_{s+1}, \ldots, u^*_T$ that yield the maximum value $J_s(x)$ in (3) do depend on x. In particular, the first control does so, and determining the functions $J_s(x)$ defined in (3) requires finding optimal closed-loop controls $u^*_s(x)$, for $s = 0, 1, \ldots, T$.

Let us denote by $\bar{u}_t = u_t(x_t)$ the control values (numbers) generated by this particular sequence of states $\{x_t\}$. Next insert these numbers $\bar{u}_t$ into the difference equation:

$$x_{t+1} = g_t(x_t, \bar{u}_t),\ x_0 \ \text{given} \tag{$**$}$$

This, obviously, has exactly the same solution as equation $(*)$.

Hence, we get the same result whether we insert the closed-loop controls $u_t^*(x)$ or the equivalent open-loop controls $\bar{u}_t$. In fact, once we have used the closed-loop controls to calculate the equivalent open-loop controls, it would seem that we can forget about the former. It may nevertheless be useful not to forget entirely the form of each closed-loop control. For suppose that at some time τ, there is a disturbance to the state x_τ^* obtained from the difference equation, which has the effect of changing the state to $\hat{x}_\tau$. Then $u_\tau^*(\hat{x}_\tau)$ still gives the optimal control to be used at that time, provided we assume that no further disturbances will occur.

EXAMPLE 5: Consider the problem of an investor with planning horizon T and initial wealth $x_0 > 0$. Let x_t denote the value of the investor's assets at time t, and let u_t be his consumption. Suppose that the interest rate on this assets is $i_t \geq 0$ at time t, so that so that $x_{t+1} = (1 + i_t)(x_t - u_t)$.

The utility associated with a level of consumption u at any one period is $u^{1-\gamma}$, where $\gamma \in (0, 1)$, and the investor wants to maximize the present value of the flow of utility from consumption. Define $\beta = 1/(1 + r)$, where r is the rate of discount. The investor's problem is, thus,

$$\max \sum_{t=0}^{T} \beta^t u_t^{1-\gamma} \ \text{subject to} \ x_{t+1} = a_t(x_t - u_t),\ u_t \in [0, x_t],$$

where, for simplicity of notation, $a_t = 1 + i_t$.

Solution: We apply Theorem 1, with the control region $U_t(x) = [0, x]$ and $f_t(x, u) = \beta^t u^{1-\gamma}$ for $t = 0, 1, \ldots, T$. By definition,

$$J_T(x) = \max_{u \in [0,x]} \beta^T u^{1-\gamma} = \beta^T x^{1-\gamma}, \tag{i}$$

and $u_T^*(x) = x$ is optimal. Moreover, Equation (4) yields

$$J_s(x) = \max_{u \in [0,x]} \left\{ \beta^s u^{1-\gamma} + J_{s+1}[a_s(x - u)] \right\}. \tag{ii}$$

In particular, (i) gives $J_T(a_{T-1}(x - u)) = \beta^T a_{T-1}^{1-\gamma}(x - u)^{1-\gamma}$, so

$$J_{T-1}(x) = \beta^{T-1} \max_{u \in [0,x]} [u^{1-\gamma} + \beta a_{T-1}^{1-\gamma}(x - u)^{1-\gamma}]. \tag{iii}$$

Let $h(u) = u^{1-\gamma} + c^\gamma(x - u)^{1-\gamma}$ denote the maximand in (iii), where $c^\gamma = \beta a_{T-1}^{1-\gamma}$. Then, $h'(u) = (1-\gamma)u^{-\gamma} - (1-\gamma)c^\gamma(x-u)^{-\gamma} = 0$ when $u^{-\gamma} = c^\gamma(x-u)^{-\gamma}$ and so $u = (x-u)/c$.

Because $\gamma \in (0,1)$ and $c^\gamma > 0$, the function h is easily seen to be concave, so the critical value of u does maximize $h(u)$. This implies that

$$u^*_{T-1}(x) = \frac{x}{w}, \text{ where } w = 1 + c = 1 + (\beta a^{1-\gamma}_{T-1})^{1/\gamma} = C^{1/\gamma}_{T-1}. \tag{iv}$$

for a suitably defined constant C_{T-1}. Then, because $\beta a^{1-\gamma}_{T-1} = c^\gamma = (w-1)^\gamma$, choosing the value x/w of u_{T-1} gives

$$\begin{aligned} h(x/w) &= x^{1-\gamma}w^{\gamma-1} + (w-1)^\gamma[x(1-w^{-1})]^{1-\gamma} \\ &= x^{1-\gamma}[w^{\gamma-1} + (w-1)^\gamma(w-1)^{1-\gamma}/w^{1-\gamma}] \\ &= x^{1-\gamma}w^\gamma \\ &= x^{1-\gamma}C_{T-1} \end{aligned}$$

Hence, by (iv), $J_{T-1}(x) = \beta^{T-1}C_{T-1}x^{1-\gamma}$.[11] Next, substitute $s = T-2$ in (ii) to get:

$$J_{T-2}(x) = \beta^{T-2} \max_{u \in [0,x]} [u^{1-\gamma} + \beta C_{T-1} a^{1-\gamma}_{T-2}(x-u)^{1-\gamma}].$$

Comparing this function with (iii), from (iv) we see that the maximum value is at $u^*_{T-2}(x) = x/C^{1/\gamma}_{T-2}$, where $C^{1/\gamma}_{T-2} = 1 + (\beta C_{T-1} a^{1-\gamma}_{T-2})^{1/\gamma}$, so that $J_{T-2}(x) = \beta^{T-2}C_{T-2}x^{1-\gamma}$.

We can go backwards repeatedly in this way and, for every t, obtain $J_t(x) = \beta^t C_t x^{1-\gamma}$. From (i), we can let $C_T = 1$, while C_t for $t < T$ is determined by backward recursion using the first-order difference equation

$$C^{1/\gamma}_t = 1 + (\beta C_{t+1} a^{1-\gamma}_t)^{1/\gamma} = 1 + (\beta a^{1-\gamma}_t)^{1/\gamma} C^{1/\gamma}_{t+1} \tag{v}$$

that is linear in $C^{1/\gamma}_t$. The optimal control is $u^*_t(x) = x/C^{1/\gamma}_t$, for all $t = 0, 1, \ldots, T-1$, while we find the optimal path by successively inserting u^*_t, into the difference equation for x_{t+1}.

We can obtain an explicit solution in the special case when $a_t = a$ for all t. Then (v) reduces to $C^{1/\gamma}_{t+1} - C^{1/\gamma}_t/\omega = -1/\omega$, where $\omega = (\beta a^{1-\gamma})^{1/\gamma}$. This is a first-order linear difference equation with constant coefficients. Using $C_T = 1$, and solving the equation for $C^{1/\gamma}_t$, we obtain

$$C^{1/\gamma}_t = \omega^{T-t} + \frac{1-\omega^{T-t}}{1-\omega}$$

for $t = T, T-1, \ldots, 0$. □

[11] Notice that J_{T-1} has the same functional form as J_T.

Problems:

(1) Consider the problem

$$\max \sum_{t=0}^{2} \left[1 - (x_t^2 + 2u_t^2)\right], \text{ subject to } x_{t+1} = x_t - u_t; x_0 = 5 \qquad (*)$$

where u_t can take any value in $\mathbb{R}$.

(a) Use Theorem 1 to solve the problem.

(b) Use the difference equation in $(*)$ to compute x_1 and x_2 in terms of u_0 and u_1, and find the sum in $(*)$ as a function S of u_0, u_1, and u_2. Next, maximize this function.

(2) Consider Problem $(*)$ in Example 1. Suppose that the utility function is $W_t(c_t) = (1+r)^{-t}\sqrt{c_t}$, and that $x_0 > 0$.[12] Compute J_s and u_s^*, for $s = T, T-1, T-2$.

(3) Consider Problem $(*)$ in Example 1. Suppose that $W_t(c_t) = (1+r)^{-t}c_t$ and $x_0 > 0$. Compute J_T, u_T^*, J_{T-1}, and u_{T-1}^*, for $x \geq 0$. Prove that there exist constants P_s, which may depend on ρ and r, such that $J_s(x) = P_s x$ for $s = 0, 1, \ldots, T$. Find J_0 and u_s^*, for $s = 0, \ldots, T$.

(4) Consider the problem

$$\max \sum_{t=0}^{T} (3 - u_t)x_t^2 \text{ subject to } x_{t+1} = u_t x_t,\ u_t \in [0,1];\ x_0 \text{ given}$$

(a) Compute the value functions J_T, J_{T-1}, and J_{T-2}, and the corresponding control functions, u_T^*, u_{T-1}^*, and u_{T-2}^*.

(b) Find an expression for $J_{T-n}(x)$, for $n = 0, 1, \ldots, T$, and the corresponding optimal controls.

(5) Given $x_0 > 0$, solve the problem

$$\max \left[\sum_{t=0}^{T-1} \left(-\tfrac{2}{3}u_t\right) + \ln x_T\right] \text{ subject to } x_{t+1} = x_t(1+u_t),\ u_t \in [0,1]$$

(6) Consider the problem

$$\max \sum_{t=0}^{T} (x_t - u_t^2) \text{ subject to } x_{t+1} = 2(x_t + u_t);\ x_0 = 0$$

(a) Write down its fundamental equations.

[12] As in Example 4, r denotes the rate of discount.

(b) Prove that the value function for the problem is given by

$$J_{T-n}(x) = (2^{n+1} - 1)x + \sum_{j=0}^{n} (2^j - 1)^2$$

for each $n = 0, 1, \dots, T$.

(c) Find optimal controls $u_t = u_t^*$ and the maximum value $V = J_0(0)$.

(7) Given positive constants α and γ, consider the problem

$$\max \left[\sum_{t=0}^{T-1} (-e^{-\gamma u_t}) - \alpha e^{-\gamma x_T} \right] \text{ subject to } x_{t+1} = 2x_t - u_t;\ x_0 \text{ given}$$

where u_t can take any value in $\mathbb{R}$.

(a) Compute J_T, J_{T-1}, and J_{T-2}.

(b) Prove that J_t can be written in the form $J_t(x) = -\alpha_t e^{-\gamma x}$, and find a difference equation for α_t.

(8) Consider the special case of Problem 2 where $r = 0$.

(a) Compute J_T, J_{T-1}, and J_{T-2}.[13]

(b) Show that the optimal control function is

$$u_s(x) = \frac{1}{1 + \rho + \rho^2 + \cdots + \rho^{T-s}},$$

and find the corresponding $J_s(x)$, $s = 1, 2, \dots, T$.

12.2 The Euler Equation

The economics literature sometimes considers the following formulation of the basic dynamic programming problem, without any explicit control variable (e.g. Stokey et al. (1989)):

$$\max \sum_{t=0}^{T} F_t(x_t, x_{t+1}), \text{ subject to } x_t \in \mathbb{R};\ x_0 \text{ given.} \tag{8}$$

That is, instead of making explicit a control, in this formulation the instantaneous reward $F_t(x_t, x_{t+1})$ at time t depends on t and on the values of the state variable at adjacent times t and $t + 1$.

If we define $u_t = x_{t+1}$, then (8) becomes a standard dynamic programming problem with $U = \mathbb{R}$. On the other hand, the dynamic optimization Problem (2) can usually be

[13] *Hint:* Consider the function $[\sqrt{u} + A\sqrt{1-u}]$, defined for $u \in [0, 1]$. Prove that this function is maximized at $u = 1/(1 + A^2)$, where its value is $\sqrt{1 + A^2}$.

formulated as a problem of the type (8).[14] To see this, let u_t be a value of $u \in U$ that maximizes $f_t(x_t, u)$ subject to the constraint that $g_t(x_t, u) = x_{t+1}$, and define $F_t(x_t, x_{t+1})$ as the maximum value of this problem. That is, define

$$F_t(x_t, x_{t+1}) = \max_u f_t(x_t, u) \text{ subject to } x_{t+1} = g_t(x_t, u),\ u \in U. \tag{9}$$

Now, let $(x_0^*, \dots, x_{T+1}^*)$ be an optimal solution of problem (8).[15] Then, $(x_1^*, \dots, x_{T+1}^*)$ is a maximum point for the objective function

$$S(x_1, \dots, x_{T+1}) = \sum_{t=0}^{T} F_t(x_t, x_{t+1}),$$

and by the usual first-order condition we must have $S_t'(x_1^*, \dots, x_{T+1}^*) = 0$ for $t = 1, \dots, T+1$. Hence, $(x_1^*, \dots, x_{T+1}^*)$ must satisfy the **Euler equations**:

$$\frac{\partial F_t}{\partial x_t}(x_t, x_{t+1}) + \frac{\partial F_{t-1}}{\partial x_t}(x_{t-1}, x_t) = 0, \quad \text{for } t = 1, \dots, T, \tag{10}$$

and

$$\frac{\partial F_T}{\partial x_{T+1}}(x_T, x_{T+1}) = 0. \tag{11}$$

Note that (10) is a second-order difference equation analogous to the Euler equation in the classical calculus of variations (see Section 8.2).[16]

EXAMPLE 6: Consider the problem

$$\max \left(\sum_{t=0}^{T-1} \ln c_t + \ln x_T \right) \text{ subject to } x_{t+1} = \alpha(x_t - c_t); x_0 \text{ given.}$$

Here x_t is wealth at time t, while c_t is the amount subtracted for consumption, and the remaining amount $x_t - c_t$ is deposited in an account and increases to $x_{t+1} = \alpha(x_t - c_t)$ at time $t+1$, where $\alpha > 1$.

Formulate the problem without explicit control variables, and use the Euler equation to solve it.

Solution: Define $\beta = 1/\alpha$. Because $c_t = x_t - \beta x_{t+1}$, the formulation without explicit control variables is

$$\max \left[\sum_{t=0}^{T-1} \ln(x_t - \beta x_{t+1}) + \ln x_T \right]. \tag{$*$}$$

[14] If, in particular, for every choice of x_t and x_{t+1} the equation $x_{t+1} = g_t(x_t, u_t)$ has a unique solution $u_t \in U$, denote this solution by $u_t = \varphi_t(x_t, x_{t+1})$. Now, define $F_t(x_t, x_{t+1}) = f_t(x_t, \varphi_t(x_t, x_{t+1}))$ for $t < T$, and $F_T(x_T, x_{T+1}) = \max_{u \in U} f_T(x_T, u)$. Then, Problem (2) becomes precisely the same as Problem (8).

[15] Remember that x_0 is given, and denote $x_0^* = x_0$.

[16] Note carefully that the partial derivatives in (10) are evaluated at different pairs. Note also that if x_{T+1} does not appear explicitly in $F_T(x_T, x_{T+1})$, Equation (11) becomes trivial.

For $t = T$, the Euler equation (10) is:[17]

$$\frac{\partial F_T}{\partial x_T}(x_T, x_{T+1}) + \frac{\partial F_{T-1}}{\partial x_T}(x_{T-1}, x_T) = 0.$$

With $F_T(x_T, x_{T+1}) = \ln x_T$ and $F_{T-1}(x_{T-1}, x_T) = \ln(x_{T-1} - \beta x_T)$, the Euler equation reduces to

$$\frac{1}{x_T} - \frac{\beta}{x_{T-1} - \beta x_T} = 0,$$

so $x_{T-1} = 2\beta x_T$.

For $t = 1, 2, \ldots, T-1$, with $F_t(x_t, x_{T+1}) = \ln(x_t - \beta x_{T+1})$, (10) gives

$$\frac{1}{x_t - \beta x_{t+1}} - \frac{\beta}{x_{t-1} - \beta x_t} = 0.$$

Solving this for x_{t-1} gives the (reverse) second-order difference equation $x_{t-1} = 2\beta x_t - \beta^2 x_{t+1}$. In particular, for $t = T-2$ this gives $x_{T-2} = 2\beta x_{T-1} - \beta^2 x_T = 4\beta^2 x_T - \beta^2 x_T = 3\beta^2 x_T$. More generally, given x_T and $x_{T-1} = 2\beta x_T$, one can show that $x_t = (T+1- \; t)\beta^{T-t} x_T$, by backward induction. This implies that $x_0 = (T+1)\beta^T x_T$, so $x_T = x_0 \beta^{-T}/(T+1)$. We thus conclude that the optimal solution of the problem is

$$x_t^* = \frac{T+1-t}{T+1}\beta^{-t} x_0, \quad c_t^* = x_t^* - \beta x_{t+1}^* = \frac{\beta^{-t} x_0}{T+1}.$$

We see that optimal consumption is steadily decreasing as t increases. □

A general solution procedure for problem (8) is similar to that used in Section 12.1.[18] It uses the Euler equations (10) and (11) recursively backwards in the following manner:[19]

---

[17]One might prefer to equate the partial derivatives of the maximand in (∗) to 0 directly, rather than introducing the function F. In particular, equating the partial derivative with respect to x_T to 0 yields

$$-\frac{\beta}{x_{T-1} - \beta x_T} + \frac{1}{x_T} = 0$$

again; equating each partial derivative with respect to x_t to 0 yields

$$-\frac{\beta}{x_{t-1} - \beta x_t} + \frac{1}{x_t - \beta x_{t+1}} = 0,$$

for $t = 1, 2, \ldots, T-1$.

[18]In Example 6 above, we used a different approach.

[19]In terms of the equations, the procedure is: first, solve X_{T+1} from

$$\frac{\partial F_T}{\partial x_{T+1}}(x_T, x_{T+1}) = 0;$$

second, solve x_T from

$$\frac{\partial F_T}{\partial x_T}[x_T, x_{T+1}^*(x_T)] + \frac{\partial F_{T-1}}{\partial x_T}(x_{T-1}, x_T) = 0;$$

(1) Use (11) to find x_{T+1} as a function of x_T, denoted by $x^*_{T+1}(x_T)$.

(2) Insert $x^*_{T+1}(x_T)$ for x_{T+1} in (10) for $t = T$, and use this equation to find x_T as a function $x^*_T(x_{T-1})$ of x_{T-1}.

(3) Insert $x^*_T(x_{T-1})$ for x_T in in (10) for $t = T - 1$ and use this equation to find x_{T-1} as a function $x^*_{T-1}(x_{T-2})$ of x_{T-2}.

(4) Continue this backward recursion until $x^*_1(x_0)$ has been found.

(5) Since x_0 is given, one can work forward to determine first $x_1 = x^*_1(x_0)$, then $x_2 = x^*_2(x_1)$, and so on..

Problems:

(1) Consider Problem 12.1.1.

   (a) Transform this problem to the form (8).

   (b) Derive the corresponding Euler equation, and find its solution. Compare with the answer to Problem 12.1.1.

(2) Consider the problem in Example 12.1.4.

   (a) Transform this problem to the form (8).

   (b) Derive the corresponding Euler equation, and find its solution. Compare with the answer to in Example 12.1.4.

12.3 Infinite Horizon

Economists often study dynamic optimization problems over an infinite horizon. This avoids the need to specify what happens after a finite horizon is reached. It also avoids having the horizon as an extra exogenous variable that features in the solution.

This section considers how dynamic programming methods can be used to study the following infinite horizon version of the problem set out in (2):

$$\max \sum_{t=0}^{\infty} \beta^t f(x_t, u_t) \text{ subject to } x_{t+1} = g(x_t, u_t), u_t \in U(x); x_0 \text{ given,} \tag{12}$$

where $U \subseteq \mathbb{R}$ is given. Here, f and g are given functions, independent of t, and there is a constant discount factor $\beta \in (0, 1)$.[20] Note that, apart from replacing the horizon T by ∞

third, solve x_{T-1} from

$$\frac{\partial F_{T-1}}{\partial x_{T-1}}[x_{T-1}, x^*_T(x_{T-1})] + \frac{\partial F_{T-2}}{\partial x_{T-1}}(x_{T-2}, x_{T-1}) = 0;$$

and so on.

[20] As pointed out in the case of finite horizon, the same theory applies without change when x_t, u_t, and g are vector functions. Having $\beta \in (0, 1)$ is essential for the subsequent analysis of the problem in this section.

as the upper limit of the sum, the two functions $f_t(x_t, u_t)$ and $g_t(x_t, u_t)$ in (2) have been replaced by $\beta^t f(x_t, u_t)$ and $g(x_t, u_t)$ respectively. Because neither the new function f nor g depends explicitly on t, Problem (12) is called **autonomous** or **stationary**.

The sequence pair $(\{x_t\}, \{u_t\})$ is called **admissible** provided that each control satisfies $u_t \in U(x_t)$, the initial state x_0 has the given value, and the difference equation in the problem is satisfied for all t.

For simplicity, we begin by assuming that f satisfies the following **boundedness condition**:[21] there exist numbers M_1 and M_2 such that

$$M_1 \leq f(x, u) \leq M_2 \text{ for all } (x, u) \text{ with } u \in U(x) \tag{13}$$

Because $0 < \beta < 1$, the sum in (12) will then always converge.

For any given starting time s, with $s = 0, 1, 2, \ldots$, and any given state x at that time, take any control sequence $\vec{u}_s = \{u_s, u_{s+1}, \ldots\}$, where $u_t \in U$ for all $t \geq s$. The successive states generated by this control sequence are found by letting $x_{t+1} = g(x_t, u_t)$, with $x_s = x$. With this notation, the discounted sum of the infinite utility (or benefit) sequence that is obtained from applying the control sequence $\vec{u}_s$, starting from state x at time s, is

$$V_s(x, \vec{u}_s) = \sum_{t=s}^{\infty} \beta^t f(x_t, u_t) \tag{14}$$

It is convenient to also define the value of that sum discounted back only to time s, namely

$$V^s(x, \vec{u}_s) = \sum_{t=s}^{\infty} \beta^{t-s} f(x_t, u_t) \tag{15}$$

It is immediate that $V_s(x, \vec{u}_s) = \beta^s V^s(x, \vec{u}_s)$. Now let

$$J^s(x) = \max_{\vec{u}_s} V^s(x, \vec{u}_s) \tag{16}$$

where for each s the maximum is taken over all sequences $\vec{u}_s = \{u_s, u_{s+1}, \ldots\}$ with $u_{s+k} \in U(x_{s+k})$.[22] If we define $J_s(x) = \beta^s J^s(x)$, we have the maximum total discounted utility (or benefit) that can be obtained over all the periods $t \geq s$, given that the system starts in state x at time s. It is immediate that

$$J_s(x) = \max_{\vec{u}_s} V_s(x, \vec{u}_s) \tag{17}$$

Note next that function $J^s(x)$ satisfies the following important property.

---

[21] It suffices to assume that this condition holds for all x in $X(x_0) = \cup_{t=0}^{\infty} X_t(x_0)$, where $X_t(x_0)$ is defined in 12.1.2.

[22] The existence of this maximum is discussed later in Note (REF?). Note that function $J_s(x)$ need only be defined on $X_s(x_0)$.

LEMMA 1. *For each period $s = 0, \ldots,$ let J^s be defined as in (17). For any s,*

$$J^0(x) = J^s(x) \tag{18}$$

The intuition of this lemma is that, because the problem is autonomous and we start in the same state x, the future looks exactly the same at either time 0 or time s.[23] In other words, finding either $J^s(x) = \max_{\vec{u}_s} V^s(x, \vec{u}_s)$ or $J^0(x) = \max_{\vec{u}_0} V^0(x, \vec{u}_0)$ requires solving essentially the same optimization problem, which therefore gives the same maximum value in each case. Importantly, the definition of function J_s and Equation (18) together imply that

$$J_s(x) = \beta^s J^0(x) \tag{19}$$

Now, define

$$J(x) = J_0(x) = J^0(x) \tag{20}$$

We call function $J(x)$ the **optimal value function**, or, more simply, the **value function** for problem (12). This function contains all the information one needs, for (19) implies that if we know $J(x)$, then we know $J_s(x)$ for all s. Moreover, the main result in this section is the following property of the value function.

THEOREM 2 (Fundamental Equation of Dynamic Programming). *The value function of Problem (12), $J(x)$, satisfies the equation*

$$J(x) = \max_{u \in U(x)} \{f(x, u) + \beta J[g(x, u)]\}. \qquad \text{(The Bellman equation)}$$

A rough argument for Theorem 2 resembles the argument for Theorem 1: Suppose we start in state x at time 0. If we choose the control u, the immediate reward is $\beta^0 f(x, u) = f(x, u)$, and at time 1 we move to state $x_1 = g(x, u)$. Choosing an optimal control sequence from time 1 on gives a total reward over all subsequent periods that equals $J_1(g(x, u)) = \beta J(g(x, u))$. Hence, the best choice of u at 0 is one that maximizes the sum $f(x, u) + \beta J(g(x, u))$. The maximum of this sum is therefore $J(x)$.

[23] Formally, let x be any fixed state and consider any policy sequence $\vec{u}_s = \{u_s, u_{s+1}, \ldots\}$ that starts at time s, and define the corresponding sequence $\vec{u}_s^0 = \{u_0^0, u_1^0, \ldots\}$ shifted earlier so that it starts at time 0 instead of at time s: let $u_t^0 = u_{s+t}$ for all t. Then, given the same starting state x, if x_t and x_t^0 denote the states reached at time t by following $\vec{u}_s$ and $\vec{u}_s^0$ starting at times s and 0 respectively, it follows by construction that $x_t^0 = x_{s+t}$ and, hence, that $f(x_t^0, u_t^0) = f(x_{s+t}, u_{s+t})$. It follows from (16) that $V^0(x, \vec{u}_s^0) = V^s(x, \vec{u}_s)$. But every shifted admissible policy $\vec{u}_s^0$ is also admissible at time 0, so we can use (17) to conclude that

$$J^0(x) = \max_{\vec{u}_0} V^0(x, \vec{u}_0) \geq V^0(x, \vec{u}_s^0) = \max_{\vec{u}_s} V^s(x, \vec{u}_s) = J^s(x).$$

By an identical argument, one can conclude that $J^0(x) \leq J^s(x)$, and hence that $J^0(x) = J^s(x)$.

EXAMPLE 7: Consider the infinite horizon analogue of Example 5, in the case when the interest rate is constant, so that $a_t = a$ for all t. Introduce a new control v defined by $u = vx$, so that v represents the proportion of wealth x that is spent in the current period. Under this change the former constraint $u \in (0, x)$ is replaced by $v \in (0, 1)$, and the problem becomes

$$\max \sum_{t=0}^{\infty} \beta^t (v_t x_t)^{1-\gamma} \text{ subject to } x_{t+1} = a(1 - v_t)x_t,\ v_t \in (0, 1). \tag{i}$$

where a and x_0 are positive constants, $\beta \in (0, 1)$, $\gamma \in (0, 1)$, and $\beta a^{1-\gamma} < 1$.[24]

Solution: In the notation of Problem (12), we have $f(x, v) = (vx)^{1-\gamma}$ and $g(x, v) = a(1 - v)x$. Thee Bellman equation therefore yields

$$J(x) = \max_{v\in(0,1)} \{(vx)^{1-\gamma} + \beta J[a(1 - v)x]\}. \tag{ii}$$

In the problem in Example 5, the value function was proportional to $x^{1-\gamma}$, so a reasonable guess in the present case is that $J(x) = kx^{1-\gamma}$ for some positive constant k. We try this as a solution, with the constant k as the only unknown. Then, after cancelling the factor $x^{1-\gamma}$, Equation (ii) reduces to the equality

$$k = \max_{v\in(0,1)} \varphi(v), \tag{iii}$$

where

$$\varphi(v) = v^{1-\gamma} + \beta k a^{1-\gamma}(1 - v)^{1-\gamma}$$

is defined on the interval $[0, 1]$. Note that $\varphi(v)$ is the sum of two functions that are concave in v. A helpful trick is to define the new constant $\rho > 0$ so that $\beta a^{1-\gamma} = \rho^\gamma$, and therefore $\varphi(v) = v^{1-\gamma} + k\rho^\gamma(1 - v)^{1-\gamma}$. The first-order condition for maximizing φ is, then,

$$\varphi'(v) = (1 - \gamma)v^{-\gamma} - (1 - \gamma)k\rho^\gamma(1 - v)^{-\gamma} = 0,$$

implying that $v^{-\gamma} = k\rho^\gamma(1 - v)^{-\gamma}$. Raising each side to the power $-1/\gamma$ and then solving for v, we see that the maximum of φ is attained at

$$v = \frac{1}{1 + \rho k^{1/\gamma}}, \text{ where } \rho = (\beta a^{1-\gamma})^{1/\gamma}. \tag{iv}$$

Now, Equation (iii) implies that k satisfies the equation

$$k = \frac{1}{(1 + \rho k^{1/\gamma})^{1-\gamma}} + k\rho^\gamma \frac{\rho^{1-\gamma} k^{(1-\gamma)/\gamma}}{(1 + \rho k^{1/\gamma})^{1-\gamma}} = (1 + \rho k^{1/\gamma})^\gamma$$

[24] Because the horizon is infinite, we may think of x_t as the assets of some institution like a university or a government that suffers from "immortality illusion" and so regards itself as timeless.

Raising each side to the power $1/\gamma$ and solving for $k^{1/\gamma}$ yields $k^{1/\gamma} = 1/(1-\rho)$, or $k = (1-\rho)^{-\gamma}$. Inserting this into (iv) gives $v = 1-\rho$, so ρ is the constant fraction of current assets that are saved in each period. Because $J(x) = kx^{1-\gamma}$, we have

$$J(x) = (1-\rho)^{-\gamma}x^{1-\gamma}, \text{ with } v = 1-\rho, \text{ where } \rho = (\beta a^{1-\gamma})^{1/\gamma} \tag{v}$$

Note that ρ increases with the discount factor β and with the return a to saving, as an economist would expect.[25] □

Whenever we wrote "max" above, it was implicitly assumed that the maximum exists. Of course, without further conditions on the system, this need not be true. Under the boundedness condition (13), the same assumptions as in the finite horizon case, namely that f and g are continuous and U is compact, do ensure that the maximum in (17) and in the right-hand side of the Bellman equation both exist.

Many economic applications, however, do not satisfy the boundedness condition, so it is important to us investigate what happens when we use the supremum instead of the maximum in Equation (17), as well as when the set $U(x)$ depends on x. In fact, suppose that $\sum_{t=0}^{\infty} \beta^t f(x_t, u_t)$ always exists.[26]. Then, $J_0(x_0) = \sup_{\vec{u}_0} V_0(x_0, \vec{u}_0)$ must exist. By the result (A.4.7) on iterated suprema, we have

$$\begin{aligned} J_0(x_0) &= \sup_{\vec{u}_0} \sum_{t=0}^{\infty} \beta^t f(x_t, u_t) \\ &= \sup_{u_0 \in U(x)} [f(x_0, u_0) + \sup_{\vec{u}_1} \sum_{t=1}^{\infty} \beta^t f(x_t, u_t)] \\ &= \sup_{u_0 \in U(x)} [f(x_0, u_0) + J_1(g(x_0, u_0))] \\ &= \sup_{u_0 \in U(x)} [f(x_0, u_0) + \beta J_0(g(x_0, u_0))]. \end{aligned}$$

[25] In this example, the boundedness condition (13) is not valid without a simple transformation. Note that $a^t x_0$ is the maximum wealth the consumer could have accumulated by time t by spending nothing (i.e. if $v_s = 0$ for $s \leq t-1$. Now define the modified state variable $y_t = x_t/(x_0 a^t)$, which is the proportion of this maximum wealth that remains. Obviously $y_0 = 1$, and y_t satisfies the difference equation $y_{t+1} = (1-v_t)y_t$, so $1 \geq y_1 \cdots y_t \geq y_{t+1} \geq \cdots \geq 0$. The new objective function is

$$\sum_{t=0}^{\infty} \hat{\beta}^t (x_0 v_t y_t)^{1-\gamma} = (x_0)^{1-\gamma} \sum_{t=0}^{\infty} \hat{\beta}^t (v_t y_t)^{1-\gamma}$$

where $\hat{\beta} = \beta a^{1-\gamma}$ and so $0 < \hat{\beta} < 1$. Using the results above for β and a replaced by $\hat{\beta}$ and 1, the maximum value $\hat{J}(y)$ of the sum is $\hat{J}(y) = (1-\hat{\rho})^{-\gamma} y^{1-\gamma}$, with $v = 1-\hat{\rho}$ and $\hat{\rho} = \hat{\beta}^{1/\gamma} = \rho$. The transformed problem satisfies the restricted boundedness condition in footnote 21, because the modified state y_t remains within the interval $[0,1]$ for all t, and so $0 \leq (x_0 y_t v_t)^{1-\gamma} \leq x_0^{1-\gamma}$ for all t and for all v_t in $[0,1]$. This implies that the control v defined in (iv) really is optimal and the transformed problem is solved. So is the original problem, of course.

[26] albeit possibly with an infinite value

So the modification

$$J(x) = \sup_{u \in U(x)} \{f(x,u) + \beta J[g(x,u)]\} \tag{21}$$

of the Bellman equation still holds even if no maximum exists.

We call the Bellman equation a "functional equation" because the unknown is the function J that appears on both sides. A natural question to ask is whether this equation has a unique solution. Importantly, *under the boundedness condition* (13), *together with the assumptions that the maximum in the right-hand side of the Bellman equation is attained and that* $0 < \beta < 1$, the equation always has one and only one bounded solution $\hat{J}$, which must therefore be the optimal value function for the problem. The value $u(x)$ of the control $u \in U$ that maximizes the right-hand side of the Bellman equation is the *optimal control*, which is therefore independent of t.

In general, it is difficult to use the Bellman equation to find $J(x)$. The reason is that the maximand in its right-hand side involves the unknown function J. Next, let us use the contraction mapping theorem 14.3.1 to prove that Equation (21) has a unique solution. Define the operator T on the domain $\mathbb{B}$ of all bounded functions $I(x)$ so that

$$T(I)(x) = \sup_{u \in U(x)} \{f(x,u) + \beta I[g(x,u)]\} \tag{**}$$

for all I and all x. As in Section 14.3, the distance between any two bounded functions $\tilde{J}$ and $\bar{J}$ is defined as $d(\tilde{J}, \bar{J}) = \sup_z |\tilde{J}(z) - \bar{J}(z)|$. Then,

$$\begin{aligned} T(\tilde{J})(x) &= \sup_{u \in U(x)} \{f(x,u) + \beta \bar{J}(g(x,u)) + \beta[\tilde{J}(g(x,u)) - \bar{J}(g(x,u))]\} \\ &\leq \sup_{u \in U(x)} [f(x,u) + \beta \bar{J}(g(x,u)) + \beta d(\tilde{J}, \bar{J})] \\ &= T(\bar{J})(x) + \beta d(\tilde{J}, \bar{J}). \end{aligned}$$

Symmetrically, $T(\bar{J})(x) \leq T(\tilde{J})(x) + \beta d(\tilde{J}, \bar{J})$, so $|T(\tilde{J})(x) - T(\bar{J})(x)| \leq \beta d(\tilde{J}, \bar{J})$, implying that

$$d[T(\tilde{J}), T(\bar{J})] = \sup_x |T(\tilde{J})(x) - T(\bar{J})(x)| \leq \beta d(\tilde{J}, \bar{J}). \tag{***}$$

Because $0 < \beta < 1$, this confirms that T is a contraction mapping, so the proof is complete.

Finally, we check that any control $u = \hat{u}$ that yields a maximum in the Bellman equation (21) is optimal. To see this, let $T^{\hat{u}}$ be the operator on $\mathbb{B}$ defined by $(**)$ when $U(x)$ takes the form $\{\hat{u}(x)\}$.[27] By definition of $\hat{u}$, the unique solution J of the Bellman equation satisfies $T^{\hat{u}}(J) = J$. Because the optimal value function satisfies the Bellman equation, this unique solution equals the optimal value function. Also, because $J^{\hat{u}}$ satisfies $(*)$ for $U(x) = \{\hat{u}(x)\}$, we have $T^{\hat{u}}(J^{\hat{u}}) = J^{\hat{u}}$. But $T^{\hat{u}}$, like T itself, is a contraction mapping, so $T^{\hat{u}}(\tilde{J}) = \tilde{J}$ has a unique solution. It follows that $J = J^{\hat{u}}$. Hence the optimal value function $J(x)$ equals the criterion value obtained by following the control policy $\hat{u}$.

[27] Leaving no choice except $u = \hat{u}(x)$.

Problems:

(1) Consider the problem

$$\max \sum_{t=0}^{\infty} \beta^t(-e^{-u_t} - \tfrac{1}{2}e^{-x_t}) \text{ subject to } x_{t+1} = 2x_t - u_t,\, u_t \in \mathbb{R};\, x_0 \text{ given},$$

where $\beta \in (0,1)$. Find a constant $\alpha > 0$ such that $J(x) = -\alpha e^{-x}$ solves the Bellman equation, and show that α is unique.

(2) Consider the following problem, with $\beta \in (0,1)$:

$$\max \sum_{t=0}^{\infty} \beta^t(-\tfrac{2}{3}x_t^2 - u_t^2) \text{ subject to } x_{t+1} = x_t + u_t,\, u_t \in \mathbb{R};\, x_0 \text{ given}.$$

(a) Suppose that $J(x) = -\alpha x^2$ solves the Bellman equation. Find a quadratic equation for α. Then, find the associated value of u^*.

(b) By looking at the objective function, show that, given any starting value x_0, it is reasonable to ignore any policy that fails to satisfy both $|x_t| \leq |x_{t-1}|$ and $|u_t| \leq |x_{t-1}|$ for all t. Does footnote 21 then apply?

12.4 Stochastic Dynamic Programming

In the previous sections of this chapter, we have assumed that the state evolves according to a deterministic difference equation: given the state and the control at period t, the value of the state at period $t+1$ is completely determined. Some of the most important problems in economics, however, are not amenable to such an assumption. Instead we must allow the possibility that, even given the state and the control in period t, the state in period $t+1$ is also influenced by stochastic shocks, so that it becomes a random variable.

Following common practice in economics and statistics, we use capital letters to denote random variables, and reserve the corresponding lower case letter for values that the random variable can take; for example X_t will denote the state at date t, whereas x_t denotes a realization of X_t. For simplicity, we concentrate on the one-dimensional case, so we assume that X_t takes values on $\mathbb{R}$, and that u_t is required to belong to a given subset U of $\mathbb{R}$.

We use subscripts to denote the value of a variable at a date, and superscripts to denote the historical sequence of values of the same variable up to that date. Specifically, for $t = 0, 1, \ldots, T$,

$$x^t = (x_0, x_1, x_2, \ldots x_t).$$

Unlike in the previous sections, then, the evolution of $\{X_t\}$ is not governed simply by a differential equation. Instead, we will assume that the history of realizations and controls, up to a period, determines the probability distribution of the control variable for the next period. Specifically, we consider two cases:

(1) Discrete state. We can assume that X_{t+1} is a random variable that takes values in a finite set $X \subset \mathbb{R}$. It is assumed that the probability that $X_{t+1} = x \in X$ may depend on the history of states up to period t and of the control in that period, and is given by the function $P_{t+1}(x, x^t, u_t)$.

(2) Absolutely continuous state. In this case, we assume that X_{t+1} takes values in $X \subset \mathbb{R}$, and assume that the probability that $X_{t+1} \leq x \in \mathbb{R}$ is given by

$$P_{t+1}(x, x^t, u_t) = \int_{-\infty}^{x} p_{t+1}(v, x^t, u_t)\, dv,$$

for some density function p_{t+1}. As before, the density function, and hence the probability distribution function, have (x^t, u_t) as arguments.

EXAMPLE 1: Consider first a two-stage decision problem, and assume that one wants to maximize the objective function

$$f_0(x_0, u_0) + E[f_1(X_1, u_1)], \tag{$*$}$$

where $f_t(x_t, u_t)$ denotes some instantaneous reward function and E denotes the expectation operator. The initial state x_0 is given, after which X_1 is determined randomly using the distribution $P_1(\cdot, x_0, u_0)$.

We can find the maximum by first maximizing with respect to u_1, and then with respect to u_0. When choosing u_1, we assume an arbitrary value of x_1, and simply maximize the reward $f_1(x_1, u_1)$, assuming that $X_1 = x_1$ is known before the maximization is carried out. The maximum point u_1^* is a function of the assumed state x_1. Inserting this function instead of u_1 into the objective function $(*)$, and then replacing x_1 by X_1 yields

$$f_0(x_0, u_0) + E[f_1(X_1, u_1^*(X_1))]. \tag{$**$}$$

A maximizing value of u_0 is then chosen, taking into account how this choice affects the distribution $P_1(\cdot, x_0, u_0)$.

When X_1 is uncertain, the following special case shows why it matters whether we can observe the realization x_1 of X_1 before choosing u_1. Suppose that $f_0(x_0, u_0) = 0$, $f_1(x_1, u_1) = x_1 u_1$, and X_1 takes the values 1 and -1 with probabilities 1/2. Suppose also that the control u must equal one of the two values 1 and -1. If we have to choose u_1 before observing x_1, then $E[X_1 u_1] = 0$. But if we can first observe x_1, then u_1 can depend on this observation. By choosing $u_1 = u_1^*(x_1) = x_1$, we can make $E[X_1 u_1] = 1$, which yields a higher value of the objective $(*)$. ¶

In the general problem, the random process determined by $\{P_t\}$ is to be controlled in the best possible manner by appropriate choices of the successive variables $\{u_t\}$. In all that follows we shall assume that the realization x_t of X_t can be observed before choosing u_t. The objective function is now the expectation

$$E\left[\sum_{t=0}^{T} f_t(X_t, u_t(X^t))\right]. \tag{1}$$

Here, each control u_t, for $t = 0, 1, 2, \ldots, T$ is a function $u_t(x^t)$ of the history of states up to period t. Such functions are called policies. For many stochastic optimization problems, including those studied here, this is the natural class of policies to consider in order to achieve an optimum.

The expectation in (1) is the sum $\sum_{t=0}^{T} E[f_t(X_t, u_t(X^t))]$ of the expectations of each successive term. These expectations, of course, depend on the probability distribution of each X_t. To calculate these, first recall that in the case of a discrete random variable, the probability that the events $X_1 = x_1$ and $X_2 = x_2$ occur jointly, given x_0, equals the probability that $X_2 = x_2$ occurs given $X^1 = x^1$, times the probability that $X_1 = x_1$ occurs given x_0. In general, thus, the probability of the history $x^t = (x_1, x_2, \ldots, x_t)$, is given by

$$P^t(x^t, u^{t-1}) = P_0(x_1, x_0, u_0) \cdot P_1(x_2, x^1, u_1) \cdot \ldots \cdot P_t(x_t, x^{t-1}, u_{t-1}), \tag{2}$$

given that the sequence of controls up to $t-1$ is u^{t-1}. In the continuous random variable case, the joint density $p^t(x^t, u^{t-1})$ is determined by the same formula, with each P_t in (2) replaced by p_t. Though not always necessary, we shall assume that f_t and P_{t+1} are continuous in (x^t, u^t).

The optimization problem is to find a sequence of policies $u_0^*(x_0), \ldots, u_T^*(x^T)$, that makes the objective (1) as large as possible. We now define

$$J_s(x^s) = \max E\left[\sum_{t=s}^{T} f_t(X_t, u_t(X^t)) \mid X^s = x^s\right], \tag{3}$$

where the expected total reward is maximized over all policy sequences $u_t = u_t(x^t)$, for $t = s, \ldots, T$. The expectation is taken over all possible sequences of realizations X_t of the random variables, given that the history of states up to period s is x^s. In this computation, each control $u_t(X^t)$ in the sequence must be applied when computing the sequence of successive states $(X_{s+1}, X_{s+2}, \ldots X_T)$

The central tool in solving optimization problems of the type (1) is the following **dynamic programming equation** or **optimality equation**:

$$J_s(x^s) = \max_{u_s}\{f_s(x_s, u_s) + E[J_{s+1}(X^{s+1}) \mid X^s = x^s, u_s]\}, \tag{4}$$

Moreover, when $s = T$ we have

$$J_T(x^T) = \max_{u_T} f_T(x_T, u_T). \tag{5}$$

These equations are similar to (6) and (7) in Theorem 12.1.1 for the deterministic case. The only significant difference is that (4) allows for uncertainty, by including the conditional expectation of J_{s+1}. As in the corresponding deterministic problem considered in Section 12.1, first (5) is used to find $u_T^*(x^T)$ and $J_T(x^T)$. Thereafter, (4) is used repeatedly in a backward recursion to find first $u_{T-1}^*(x^{T-1})$ and $J_{T-1}(x^{T-1})$, then $u_{T-2}^*(x^{T-2})$ and $J_{T-2}(x^{T-2})$, and so on all the way back to $u_0^*(x_0)$ and $J_0(x_0)$.

As in the deterministic case, equations (4) and (5) are, essentially, both necessary and sufficient. They are sufficient in the sense that if $u_t^*(x^t)$ maximizes the right-hand side of (4) for $t = 0, \ldots, T-1$, and also $u_T^*(x^T)$ maximizes the right-hand side of (5), then $\{u_t^*\}$ is indeed an optimal policy sequence. On the other hand, the same equations are necessary in the sense that, for every pair x^t that occurs with positive probability (or has a positive probability density when there is a continuous density function), an optimal control $u_t^*(x^t)$ must yield a maximum on the right-hand side of (4) for $s = 0, 1, \ldots, T-1$, and $u_T^*(x^T)$ one of (5).

An important special case occurs when the probability distribution of X_{t+1} depends on the control at t, and only on the last realization of the state, x_t. In this case, we can write this distribution simply as $P_{t+1}(\cdot, x_t, u_t)$, or the associated density as $p_{t+1}(\cdot, x_t, u_t)$. In this case, the optimal control of period t depends only on the value of x_t, and not on x^{t-1}, so that we can re-express $J_t(x_t)$ and $u_t(x_t)$ as a function of the realized value of X_t only. These policy functions are called **Markov policies** or **Markov controls**, to emphasize their independence from earlier values of the state variable.

A particularly simple case is when the perturbations to the state variable are independently distributed. In this case, one assumes that for $t = 0, 1, \ldots, T$ the state equation takes the form

$$X_{t+1} = g_t(x_t, u_t, \Omega_{t+1}), \tag{6}$$

with x_t given, where Ω_{t+1} is a random variable whose probability distribution does not depend on x^t, u^t or Ω^t. In this case, we can still express J_t and u_t as a function of x_t only (and not of the realized value of Ω_t), and equation (4) continues to hold. Intuitively, this is because $\Omega_t = \omega_t$ would not appear as a conditioning variable in (4). Formally, this independence can be proved by backward induction. Some examples below are of this special form.

EXAMPLE 2: Consider a stochastic version of Example 12.1.4, where at each time $t = 0, 1, \ldots, T-1$, the state variable X_t is an investor's wealth, the control variable u_t is consumption, and the certain return a_t to investment in that example is replaced by a random return Ω_{t+1}. Moreover, suppose that $\{\Omega_t\}_{t=1}^T$ is a sequence of independently distributed random variables with positive values, and that the state X_t is assumed to evolve according to the stochastic difference equation

$$X_{t+1} = \Omega_{t+1}(X_t - u_t), \quad \text{for } u_t \in [0, xt], \tag{i}$$

with x_0 given. The objective function is the obvious counterpart of that in Example 12.1.4, namely the expected sum of discounted utility, given by

$$E\left[\sum_{t=0}^{T-1} \beta^t u_t^{1-\gamma} + \beta^T A X_T^{1-\gamma}\right], \tag{ii}$$

where $\beta \in (0,1)$ is a discount factor, while $\gamma \in (0,1)$ is a taste parameter, and A is a positive constant. Thus, the problem is to maximize (ii) subject to (i). Assume that $E[\Omega_t^{1-\gamma}] < \infty$ for all t.

Solution: Here, $J_T(x^T) = \beta^T A x_T^{1-\gamma}$, exactly as in (ii) in Example 12.1.4. Because the random variables Ω_t are independently distributed, the value functions take the form $J_t(x_t)$. To find $J_{T-1}(x)$, we use the optimality equation

$$J_{T-1}(x) = \max_u \{\beta^{T-1} u^{1-\gamma} + E[\beta^{T-1} A(\Omega_T(x-u))^{1-\gamma}]\}. \tag{$*$}$$

The expectation must be calculated by using the probability distribution for Ω_T. In fact, the right-hand side of $(*)$ is of the same form as (iv) in Example 12.1.4. To make it exactly the same, define the new constant a_{T-1} so that $a_{T-1}^{1-\gamma} = E[\Omega_T^{1-\gamma}]$. With this new notation, the optimal control u_{T-1} is given by (v) in Example 12.1.4. Furthermore, equations (vii), (viii), and (ix) of that example still hold provided we define each a_t to satisfy $a_t^{1-\gamma} = E[\Omega_{T+1}^{1-\gamma}]$, for $t = 0, 1, \ldots, T-1$. One may call each at the certainty equivalent return because the solution is exactly the same as if it replaced the uncertain return described by the random variable Ω_{t+1}. Of course, each a_t depends on the taste parameter γ as well as on the distribution of Ω_t. ¶