1 Dynamic Programming

This chapter introduces the basic ideas and methods of dynamic programming and displays the restrictions on a dynamic system and the objective function that must be met for dynamic programming to be applicable. Where these restrictions are satisfied, dynamic programming provides a powerful method for studying dynamic optimization. The required restrictions permit the analyst to break what is in general a single exceedingly large dimensional optimization problem into a collection of much smaller optimization problems that can be solved sequentially. That step usually affords computational simplicity and often provides analytical insights.

The restrictions on objective functions and the dynamic system required for dynamic programming are satisfied in many formulations of private agents' investment problems. There is a class of multiagent problems (differential games), however, in which the structure of interactions among different agents' decision problems prevents one or more agents' problems from conforming to the restrictions required for dynamic programming. For these agents, therefore, optimal decisions must be computed not sequentially but simultaneously. The inapplicability of sequential methods to such decision problems is called time inconsistency and was first studied in macroeconomic contexts by Kydland and Prescott (1977) and Calvo (1978).

Although the main ideas of dynamic programming are simple, the details can involve sophisticated mathematical arguments. In this chapter things have been kept at a heuristic level of presentation, with the hope of communicating the main ideas quickly and enabling the reader to use these techniques to solve problems. More thorough presentations of the subject are listed at the end of the chapter; in particular see Bertsekas (1976); Bertsekas and Shreve (1978); Lucas, Prescott, and Stokey (forthcoming); Bellman (1957); and Chow (1981).

1.1 A General Intertemporal Problem

Consider the following general intertemporal optimization problem under certainty. Let x_t be an $(n \times 1)$ vector of *state* variables at time t, $t = 0, 1, \ldots, T+1$. Let u_t be a $(k \times 1)$ vector of *control* variables at time t, $t = 0, \ldots, T$. (The terms "state" and "control" are ambiguous in the context of the problem of this section. A precise description of them will be postponed, pending consideration of the special problem of Section 1.2, in which they are well motivated.) The problem is to choose $u_0, u_1, \ldots, u_T, x_1, \ldots, x_{T+1}$ to maximize an objective function

$$R(x_0, u_0, x_1, u_1, \ldots, x_T, u_T, x_{T+1}), \tag{1.1}$$

subject to x_0 given and subject to a system of constraints connecting the controls and the states, which we write in the implicit form

$$G(x_0, u_0, x_1, u_1, \ldots, x_T, u_T, x_{T+1}) \geq 0. \tag{1.2}$$

In (1.2) we imagine that G is a collection of $(T+1)n$ functions. We imagine that R and G are sufficiently smooth and that R is sufficiently concave to permit the method of Kuhn and Tucker to be applied. We then have a standard classical constrained-optimization problem, which can be solved by forming the following Lagrangian and maximizing with respect to $u_0, u_1, \ldots, u_T, x_1, x_2, \ldots, x_{T+1}$:

$$\begin{aligned} J = {} & R(x_0, u_0, x_1, u_1, \ldots, x_T, u_T, x_{T+1}) \\ & + \mu' G(x_0, u_0, x_1, u_1, \ldots, x_T, u_T, x_{T+1}), \end{aligned} \tag{1.3}$$

where μ is a $[(T+1)n \times 1]$ vector of Lagrange multipliers. The solution of this problem can be represented as a set of functions $u_0 = H_0(x_0)$, $u_1 = H_1(x_0), \ldots, u_T = H_T(x_0)$, with the optimal controls expressed as a function of the initial given state x_0, and a set of functions $x_1 = w_1(x_0)$, $x_2 = w_2(x_0), \ldots, x_{T+1} = w_{T+1}(x_0)$, with subsequent states expressed as a function of the initial state x_0. It is a standard feature of this problem that the optimal controls $u_0, u_1, \ldots, u_T$ as functions of x_0 must be determined simultaneously. This feature can be verified by obtaining the first-order necessary conditions for maximizing (1.3) and by studying the structure of the Jacobian matrix for the system of first-order necessary conditions. The system of first-order conditions in general fails to be recursive or block recursive, so that the optimal values for u_t and x_t are simultaneously determined.

1.2 A Recursive Problem

For dynamic problems in which the horizon T is large, it would be convenient if the problem could somehow be specialized to avoid the need to compute all of the controls simultaneously. This consideration has led to the following specialization of (1.1) and (1.2), which permits a recursive approach to the computation of the optimal controls.

We assume that $r_t(x_t, u_t)$ is a concave function and that the set $\{x_{t+1}, x_t, u_t : x_{t+1} \leq g_t(x_t, u_t),\ u_t \in R^k\}$ is convex and compact. We thus replace (1.1) and (1.2) with the problem of maximizing by choice of $(u_0, x_1, u_1, \ldots, x_{T+1})$ the function

$$r_0(x_0, u_0) + r_1(x_1, u_1) + \ldots + r_T(x_T, u_T) + W_0(x_{T+1}), \tag{1.4}$$

subject to x_0 given and the "transition" equations

$$\begin{aligned} x_1 &= g_0(x_0, u_0) \\ x_2 &= g_1(x_1, u_1) \\ &\cdot \\ &\cdot \\ &\cdot \\ x_{T+1} &= g_T(x_T, u_T). \end{aligned} \tag{1.5}$$

The function $r_t(x_t, u_t)$ is called the one-period return function at t, whereas the function $g_t(x_t, u_t)$ is called the transition function at t. The structure of the transition equations (1.5) motivates the labeling of x_t as state and u_t as control variables. The state vector x_t constitutes a complete description of the current position of the system. As far as the current and future returns $r_s(x_s, u_s)$ for $s \geq t$ are concerned, past values of u_v and x_v for $v < t$ add no information beyond that contained in x_t. This result is a consequence of the particular time separable structure of (1.4) and (1.5). The control vector u_t contains variables under the partial control of the problem solver that impinge on x_{t+1}, given x_t. In general for a given problem, the appropriate definition of the state is not unique, there being alternative ways of completely describing the current position of the system. Many of the admissible definitions of the state will include redundancies.

In (1.4) and (1.5) the functions $r_t(x_t, u_t)$, $W_0(x_{T+1})$, and $g_t(x_t, u_t)$ are assumed to be sufficiently smooth to permit the use of Lagrange's method.

Forming the Lagrangian, we have

$$L = r_0(x_0, u_0) + r_1(x_1, u_1) + \dots + r_T(x_T, u_T) + W_0(x_{T+1}) + \lambda_0'[g_0(x_0, u_0) - x_1] + \lambda_1'[g_1(x_1, u_1) - x_2] + \dots + \lambda_T'[g_T(x_T, u_T) - x_{T+1}], \tag{1.6}$$

where λ_t is an $(n \times 1)$ vector of Lagrange multipliers for $t = 0, \dots, T$ and the prime denotes transposition.

The first-order necessary conditions for this problem are

$$\frac{\partial L}{\partial u_t} = \frac{\partial r_t}{\partial u_t}(x_t, u_t) + \frac{\partial g_t(x_t, u_t)}{\partial u_t}\lambda_t = 0, \qquad t = 0, \dots, T \tag{1.7a}$$

$$\frac{\partial L}{\partial x_t} = \frac{\partial r_t(x_t, u_t)}{\partial x_t} + \frac{\partial g_t}{\partial x_t}(x_t, u_t)\lambda_t - \lambda_{t-1} = 0, \qquad t = 1, \dots, T \tag{1.7b}$$

$$\frac{\partial L}{\partial x_{T+1}} = W_0'(x_{T+1}) - \lambda_T = 0 \tag{1.7c}$$

$$x_{t+1} = g_t(x_t, u_t), \qquad t = 0, 1, \dots T. \tag{1.7d}$$

Here $\partial r_t/\partial u_t$ is a $(k \times 1)$ vector with $\partial r_t/\partial u_{ti}$ in the ith row, where u_{ti} is the element in the ith row of u_t. Also, $\partial g_t/\partial u_t$ is a $(k \times n)$ matrix with $\partial g_{ti}/\partial u_{th}$ in the ith column and hth row, where g_{ti} is the ith row of g_t and u_{th} is the hth row of u_t. Solving (1.7b) for λ_{t-1} and shifting forward one period, we have

$$\lambda_t = \frac{\partial r_{t+1}(x_{t+1}, u_{t+1})}{\partial x_{t+1}} + \frac{\partial g_{t+1}(x_{t+1}, u_{t+1})}{\partial x_{t+1}}\lambda_{t+1}.$$

Using this and (1.7c) recursively to eliminate λ_t, $t = 0, \dots, T$, from (1.7a), we obtain the following system:

$$\frac{\partial r_t}{\partial u_t}(x_t, u_t) + \frac{\partial g_t(x_t, u_t)}{\partial u_t}\left\{\frac{\partial r_{t+1}}{\partial x_{t+1}} + \frac{\partial g_{t+1}}{\partial x_{t+1}}\left[\frac{\partial r_{t+2}}{\partial x_{t+2}} + \frac{\partial g_{t+2}}{\partial x_{t+2}} \cdot \left(\frac{\partial r_{t+3}}{\partial x_{t+3}} + \frac{\partial g_{t+3}}{\partial x_{t+3}}\left\{\dots + \frac{\partial g_T}{\partial x_T}[W_0'(x_{T+1})]\right\}\right)\right]\right\} = 0 \tag{1.8a}$$

$$t = 0, \dots, T-1$$

$$x_{t+1} = g_t(x_t, u_t), \qquad t = 0, \dots, T-1 \tag{1.8b}$$

$$\frac{\partial r_T}{\partial u_T}(x_T, u_T) + \frac{\partial g_T(x_T, u_T)}{\partial u_T}W_0'(x_{T+1}) = 0 \tag{1.8c}$$

$$x_{T+1} = g_T(x_T, u_T), \tag{1.8d}$$

where in (1.8a) it is understood that g_t and r_t both have arguments (x_t, u_t).

In the special case in which $r_t(x_t, u_t)$ is quadratic, g_t is linear, and $\partial g_t/\partial x_t \equiv 0$, Equations (1.8a)–(1.8b) can be solved to yield a system of second-order difference equations in the vector x_t, subject to the initial condition that x_0 is given, and the terminal conditions (1.8c)–(1.8d). A further specialization results if the functions r_t and g_t are assumed to be time invariant so that (1.8) yields a set of time-invariant linear difference equations. In this case, the equations can be solved using methods similar to those illustrated in Sargent (1986, chap. 9). For more general specifications, however, it is useful to have an alternative method of solving the problem or at least of characterizing the solution, because nonlinear difference equations are generally very difficult to solve directly.

To motivate this method, notice the special structure of system (1.8), which is depicted in Table 1.1. The structure is special because (x_s, u_s) for $s < t$ does not appear directly in the marginal conditions and transition laws dated t and later. This fact makes it feasible to use the following "backward" recursive solution strategy.

Given x_T, the (subsystems of the) last two equations of system (1.8), namely (1.8c) and (1.8d), form a system of $(n + k)$ equations in (x_{T+1}, u_T). We solve these equations for x_{T+1} and u_T as functions of x_T, say,

$$x_{T+1} = f_T(x_T), \qquad u_T = h_T(x_T), \tag{1.9}$$

where $f_T(x_T) \equiv g_T[x_T, h_T(x_T)]$. Next, use $u_T = h_T(x_T)$ to eliminate u_T from the preceding two (subsystems of) equations in (1.8), namely (1.8a) and

Table 1.1 The structure of system (1.8)

Define the implicit functions ϕ_t^1, ϕ_t^2, $t = 0, 1, 2, \dots, T$, by

$$\phi_t^1(x_t, u_t, x_{t+1}, u_{t+1}, \dots, x_T, u_T, x_{T+1}) = \frac{\partial r_t}{\partial u_t} + \frac{\partial g_t}{\partial u_t}\left\{\frac{\partial r_{t+1}}{\partial x_{t+1}} + \frac{\partial g_{t+1}}{\partial x_{t+1}}\left(\dots + \frac{\partial g_T}{\partial x_T}\right)[W_0'(x_{T+1})]\right\} = 0$$

and $\phi_t^2(x_t, u_t, x_{t+1}) = x_{t+1} - g_t(x_t, u_t) = 0$, respectively. Then (1.8) can be represented as

$$\begin{aligned}
\phi_0^1(x_0, u_0, x_1, u_1, x_2, \dots, x_T, u_T, x_{T+1}) &= 0\\
\phi_0^2(x_0, u_0, x_1) &= 0\\
\phi_1^1(x_1, u_1, x_2, \dots, x_T, u_T, x_{T+1}) &= 0\\
\phi_1^2(x_1, u_1, x_2) &= 0\\
\phi_T^1(x_T, u_T, x_{T+1}) &= 0\\
\phi_T^2(x_T, u_T, x_{T+1}) &= 0
\end{aligned}$$

(1.8b) for $t = T-1$,

$$(1.10)\quad \frac{\partial r_{T-1}(x_{T-1}, u_{T-1})}{\partial u_{T-1}} + \frac{\partial g_{T-1}(x_{T-1}, u_{T-1})}{\partial u_{T-1}} \cdot \left[\frac{\partial r_T}{\partial x_T}(x_T, u_T) + \frac{\partial g_T}{\partial x_T}(x_T, u_T)\, W_0'(x_{T+1})\right] = 0$$

$$x_T = g_{T-1}(x_{T-1}, u_{T-1}),$$

and solve these equations for u_{T-1} and x_T each as functions of x_{T-1}:

$$(1.11)\quad x_T = f_{T-1}(x_{T-1}), \qquad u_{T-1} = h_{T-1}(x_{T-1}).$$

One can continue recursively in this way, solving for a collection of feedback rules of the form

$$(1.12)\quad u_t = h_t(x_t), \qquad t = T, T-1, T-2, \ldots, 0,$$

where $u_t = h_t(x_t)$, $x_{t+1} = f_t(x_t)$ solve the equations

$$(1.13)\quad \frac{\partial r_t}{\partial u_t}(x_t, u_t) + \frac{\partial g_t}{\partial u_t}\left\{\frac{\partial r_{t+1}}{\partial x_{t+1}} + \frac{\partial g_{t+1}}{\partial x_{t+1}}\left[\frac{\partial r_{t+2}}{\partial x_{t+2}} + \frac{\partial g_{t+2}}{\partial x_{t+2}} \cdot \left(\frac{\partial r_{t+3}}{\partial x_{t+3}} + \frac{\partial g_{t+3}}{\partial x_{t+3}}\left\{\ldots + \frac{\partial g_T}{\partial x_T}[W_0'(x_{T+1})]\right\}\right)\right]\right\} = 0,$$

and $x_{s+1} = g_s(x_s, u_s)$ for $s = t, t+1, \ldots, T$, given that $u_{s+1} = h_{s+1}(x_{s+1})$ for $s = t, t+1, \ldots, T-1$.

1.3 Bellman's Equations

The equations of system (1.13) have interpretations as the marginal conditions from the following sequence of problems. Define the value function for a one-period problem $W_1(x_T)$ by

$$(1.14)\quad W_1(x_T) = \max_{u_T}\{r_T(x_T, u_T) + W_0(x_{T+1})\},$$

subject to $x_{T+1} = g_T(x_T, u_T)$, with x_T given. We form the Lagrangian for this problem, and the first-order conditions can be expressed, after the Lagrange multiplier has been eliminated, as

$$(1.15)\quad \frac{\partial r_T}{\partial u_T}(x_T, u_T) + \frac{\partial g_T(x_T, u_T)}{\partial u_T} W_0'(x_{T+1}) = 0,$$

which precisely matches the marginal condition for u_T in (1.8). Equation (1.15) and the transition law $x_{T+1} = g_T(x_T, u_T)$ are to be solved jointly for $u_T = h_T(x_T)$. Now imagine substituting the solution $u_T = h_T(x_T)$ of (1.15) and (1.8d) into (1.14) to get

$$(1.16)\quad W_1(x_T) = r_T[x_T, h_T(x_T)] + W_0(g_T[x_T, h_T(x_T)]).$$

Formally, differentiating (1.16) gives

$$W_1'(x_T) = \left(\frac{\partial r_T}{\partial x_T} + \frac{\partial g_T}{\partial x_T} W_0'\right) + \frac{\partial h_T}{\partial x_T}\left[\frac{\partial r_T}{\partial u_T} + \frac{\partial g_T}{\partial u_T} W_0'(x_{T+1})\right],$$

where all functions dated T are evaluated at $[x_T, h_T(x_T)]$ and which by virtue of (1.15) becomes

$$(1.17)\quad W_1'(x_T) = \frac{\partial r_T}{\partial x_T}[x_T, h_T(x_T)] + \frac{\partial g_T}{\partial x_T}[x_T, h_T(x_T)] W_0'(g_T[x_T, h_T(x_T)]).$$

Because we have not shown that $\partial h_T/\partial x_T$ exists, this argument is informal or heuristic and should be regarded as only a way of remembering the correct answer. Correct arguments are given by Benveniste and Scheinkman (1979) and Lucas (1977).

Now define the value function for the two-period problem $W_2(x_{T-1})$ as

$$(1.18)\quad W_2(x_{T-1}) = \max_{u_{T-1}}\{r_{T-1}(x_{T-1}, u_{T-1}) + W_1(x_T)\},$$

subject to $x_T = g_{T-1}(x_{T-1}, u_{T-1})$, with x_{T-1} given. If we proceed as with the problem defined by (1.14), the first-order condition for the problem on the right side of (1.18) can be expressed as

$$\frac{\partial r_{T-1}}{\partial u_{T-1}}(x_{T-1}, u_{T-1}) + \frac{\partial g_{T-1}(x_{T-1}, u_{T-1})}{\partial u_{T-1}} W_1'(x_T) = 0.$$

If we use formula (1.17) for $W_1'(x_T)$, this equation becomes

$$(1.19)\quad \frac{\partial r_{T-1}}{\partial u_{T-1}}(x_{T-1}, u_{T-1}) + \frac{\partial g_{T-1}(x_{T-1}, u_{T-1})}{\partial u_{T-1}} \cdot \left(\frac{\partial r_T}{\partial x_T}[x_T, h_T(x_T)] + \frac{\partial g_T}{\partial x_T}[x_T, h_T(x_T)] W_0'(g_T[x_T, h_T(x_T)])\right) = 0.$$

This equation and the transition law $x_T = g_{T-1}(x_{T-1}, u_{T-1})$ are to be solved jointly for $u_{T-1} = h_{T-1}(x_{T-1})$, $x_T = f_{T-1}(x_{T-1})$. Again proceeding as above, we can obtain

$$(1.20)\quad W_2'(x_{T-1}) = \frac{\partial r_{T-1}}{\partial x_{T-1}}[x_{T-1}, h_{T-1}(x_{T-1})] + \frac{\partial g_{T-1}}{\partial x_{T-1}} W_1'(g_{T-1}[x_{T-1}, h_{T-1}(x_{T-1})]),$$

or, using (1.17),

$$W_2'(x_{T-1}) = \frac{\partial r_{T-1}}{\partial x_{T-1}}[x_{T-1}, h_{T-1}(x_{T-1})] + \frac{\partial g_{T-1}}{\partial x_{T-1}}[x_{T-1}, h_{T-1}(x_{T-1})] \cdot \left\{\frac{\partial r_T}{\partial x_T}[x_T, h_T(x_T)] + \frac{\partial g_T}{\partial x_T}[x_T, h_T(x_T)]\, W_0'[f_T(x_T)]\right\},$$

where x_T is evaluated at $x_T = f_{T-1}(x_{T-1}) = g_{T-1}[x_{T-1}, h_{T-1}(x_{T-1})]$. Notice that Equation (1.19) is precisely the version of the marginal condition in (1.8) for u_{T-1}.

The pattern for the recursion is now set. We iterate on the following functional equation in the value functions

$$W_{j+1}(x_{T-j}) = \max_{u_{T-j}}\{r_{T-j}(x_{T-j}, u_{T-j}) + W_j(x_{T-j+1})\}, \tag{1.21}$$

subject to $x_{T-j+1} = g_{T-j}(x_{T-j}, u_{T-j})$, x_{T-j} given. The functional equation (1.21) is a version of Bellman's equation—named after Richard Bellman (1957). The idea is to proceed recursively and to work backward, first solving the one-period problem with $j + 1 = 1$, deducing $W_1(x_T)$, then solving the two-period problem with $j + 1 = 2$, deducing the two-period value function $W_2(x_{T-1})$. The process is repeated until we obtain the $(T + 1)$-period value function $W_{T+1}(x_0)$. This procedure gives the optimal value of the problem as a function of the initial state x_0. Along the way we have calculated the optimal feedback rules $u_{T-j} = h_{T-j}(x_{T-j})$, $j = 0, 1, \ldots, T$. The preceding argument suggests that this backward recursion generates the same marginal conditions as the original problem (1.8). Indeed, the backward recursion technique always solves the original problem if a solution exists.

The derivative of the value functions obeys the recursion

$$W_{j+1}'(x_{T-j}) = \frac{\partial r_{T-j}}{\partial x_{T-j}}[x_{T-j}, h_{T-j}(x_{T-j})] + \frac{\partial g_{T-j}}{\partial x_{T-j}} W_j'(g_{T-j}[x_{T-j}, h_{T-j}(x_{T-j})]).$$

Comparing this equation with (1.7b) and (1.7c), we find that $W_j'(x_{T+1-j}) = \lambda_{T-j}$. The Lagrange multipliers λ_{T-j} in (1.6) thus give the marginal value of the state variables for the j-period problem.

The following observations supply another perspective on the recursive nature of our problem. Let us simply define the $(T + 1)$-period value function $W_{T+1}(x_0)$ by

$$W_{T+1}(x_0) = \max_{u_0, u_1, \ldots, u_T}\{r_0(x_0, u_0) + r_1(x_1, u_1) + \ldots + r_T(x_T, u_T) + W_0(x_{T+1})\}, \tag{1.22}$$

where the maximization is understood to be subject to $x_{t+1} = g_t(x_t, u_t)$, $t = 0, \ldots, T$, and x_0 given. Notice that the objective function and constraints (transition equations) have been specialized to have the key property that controls dated t influence states x_{s+1} and returns $r_s(x_s, u_s)$ for $s \geq t$ but not earlier. This key property gives the problem its recursive structure. In particular, the property makes it legitimate to cascade the maximization operator and to write (1.22) as

$$W_{T+1}(x_0) = \max_{u_0}\{r_0(x_0, u_0) + \max_{u_1}\{r_1(x_1, u_1) + \max_{u_2}\{r_2(x_2, u_2) + \ldots + \max_{u_T}\{r_T(x_T, u_T) + W_0(x_{T+1})\} \ldots \}\}\}, \tag{1.23}$$

where the maximization over u_t is understood to be subject to $x_{t+1} = g_t(x_t, u_t)$ with x_t given. Equation (1.23) indicates that the original large optimization problem on the right side of (1.22) can be broken up into $(T + 1)$ smaller problems. First, the problem in the innermost brackets is solved, the optimizer being $u_T = h_T(x_T)$ and the optimized value being $W_1(x_T)$. Then the problem in the second innermost brackets is solved for $u_{T-1} = h_{T-1}(x_{T-1})$ with optimized value $W_2(x_{T-1})$. This process of proceeding from the problems in the innermost brackets outward is equivalent to iterating on Bellman's functional equation (1.21).

The preceding argument implies that the optimal policies $u_t = h_t(x_t)$, $t = 0, \ldots, T$ have a self-enforcing character in the following sense. Consider the "remainder" of the objective function at some time $s > 0$, namely,

$$\max_{u_s, u_{s+1}, \ldots, u_T}\{r_s(x_s, u_s) + \ldots + r_T(x_T, u_T) + W_0(x_{T+1})\}, \tag{1.24}$$

subject to $x_{t+1} = g_t(x_t, u_t)$, $t = s, \ldots, T$, with x_s given. Then the solution of the maximum problem (1.24) is simply to use the remaining functions $u_s = h_s(x_s)$, $s = t, \ldots, T$ that were computed for the original problem. Furthermore, the maximized value of (1.24) is $W_{T-s+1}(x_s)$. Thus as time advances, there is no incentive to depart from the original plan. This self-enforcing character of optimal policies is known as Bellman's principle of optimality. Optimal policies that have this property are said to be time consistent. This property is special, is a consequence of the recursive character of the problem (1.4)–(1.5) and will not characterize the solutions of more general problems.

It is a feature of the solution to problem (1.4)–(1.5) that in general a different policy function $u_t = h_t(x_t)$, mapping the state at t into the control at t, is to be used at each date $t = 0, \ldots, T$. This is a consequence of two features of the problem: the fact that the horizon T is finite and the fact that the functions $r_t(x_t, u_t)$ and $g_t(x_t, u_t)$ have been permitted to depend on time in arbitrary ways. For many practical applications it is inconvenient that the policy function varies over time. One would like to discover contexts in which the same policy function is used for each period t. In the interests of achieving this objective, we now specialize problem (1.4)–(1.5) with the aim of generating conditions under which the policy functions h_j converge as $j \to -\infty$. We assume that

$$r_t(x_t, u_t) = \beta^t r(x_t, u_t), \qquad 0 < \beta < 1 \tag{1.25}$$
$$g_t(x_t, u_t) = g(x_t, u_t).$$

With this specification, Bellman's equation (1.21) becomes

$$W_{j+1}(x_{T-j}) = \max_{u_{T-j}}\{\beta^{T-j} r(x_{T-j}, u_{T-j}) + W_j(x_{T-j+1})\}.$$

Multiplying both sides by β^{j-T} gives

$$\beta^{j-T} W_{j+1}(x_{T-j}) = \max_{u_{T-j}}\{r(x_{T-j}, u_{T-j}) + \beta \cdot \beta^{j-1-T} W_j(x_{T-j+1})\}. \tag{1.25'}$$

Now define the current value function

$$V_{j+1}(x_{T-j}) = \beta^{j-T} W_{j+1}(x_{T-j}).$$

Notice that for $j = T$, we have $V_{T+1}(x_0) = W_{T+1}(x_0)$. Also notice that the current value function can be directly defined as

$$V_{j+1}(x_{T-j}) = \max_{u_{T-j}, u_{T-j+1}, \ldots, u_T}\{r(x_{T-j}, u_{T-j}) + \beta r(x_{T-j+1}, u_{T-j+1}) + \ldots + \beta^j r(x_T, u_T) + \beta^{j+1} V_0(x_{T+1})\}.$$

In terms of the current value function, (1.25′) asserts that Bellman's equation becomes

$$V_{j+1}(x_{T-j}) = \max_{u_{T-j}}\{r(x_{T-j}, u_{T-j}) + \beta V_j(x_{T-j+1})\}, \tag{1.26}$$

subject to $x_{T-j+1} = g(x_{T-j}, u_{T-j})$ and x_{T-j} given. More compactly, we can write (1.26) as

$$V_{j+1}(x) = \max_{u}\{r(x, u) + \beta V_j(\tilde{x})\}, \tag{1.27}$$

subject to $\tilde{x} = g(x, u)$, x given, where the tilde denotes next-period values. Under particular conditions, iterations on (1.27) starting from any bounded and continuous initial V_0 converge as $j \to \infty$. See Bertsekas (1976, chap. 6) or Lucas, Prescott, and Stokey (forthcoming). The argument used to prove this claim is outlined in Section A.8. In this case the limit function $V = \lim_{j\to\infty} V_j$ satisfies the following version of Bellman's equation:

$$V(x) = \max_{u}\{r(x, u) + \beta V(\tilde{x})\}, \tag{1.28}$$

where the maximization is subject to $\tilde{x} = g(x, u)$, with x given. The limiting value function V that solves (1.28) turns out to be the optimal value function for the infinite horizon problem

$$V(x_0) = \max_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t r(x_t, u_t), \tag{1.29}$$

where the maximization is subject to $x_{t+1} = g(x_t, u_t)$, with x_0 given. Problem (1.29) is a version of a discounted dynamic programming problem. Under various particular regularity conditions,[1] it turns out that (1) the functional equation (1.28) has a unique strictly concave solution; (2) this solution is approached in the limit as $j \to \infty$ by iterations on (1.26) starting from any bounded and continuous initial V_0; (3) there is a unique and time-invariant optimal policy of the form $u_t = h(x_t)$, where h is chosen to maximize the right side of (1.28); (4) off corners, the limiting value function V is differentiable with

$$V'(x) = \frac{\partial r}{\partial x}[x, h(x)] + \beta \frac{\partial g}{\partial x}[x, h(x)]\, V'(g[x, h(x)]). \tag{1.30}$$

This is a version of the formula of Benveniste and Scheinkman (1979). It is a great convenience of the specialization (1.25) of the objective function and transition functions, and also a convenience of the specification of an infinite horizon, that they imply a time-invariant policy function $u_t = h(x_t)$, for it is a routine practice in economics to seek setups in which agents use time-invariant decision rules. (Ample econometric considerations recommend or require such setups.)

The preceding results provide two methods for solving the functional

1. Alternative sets of regularity conditions work. One set of sufficient conditions is (1) r is concave and bounded, (2) the constraint set generated by g is convex and compact, that is, the set of $\{x_{t+1}, x_t, u_t : x_{t+1} \leq g(x_t, u_t)\}$ for admissible u_t is convex and compact. See Lucas (1977), and Bertsekas (1976) for further details of convergence results. See Benveniste and Scheinkman (1979) and Lucas (1977) for the results on differentiability of the value function. A proof of the uniform convergence of iterations on (1.27) is contained in Section A.7 of the Appendix.

equation (1.28). The first method is constructive and simply involves iterating on (1.26), starting from $V_0 = 0$, until V_j has converged. The second method involves guessing a solution V and verifying that it is a solution to (1.28). The second method relies on the uniqueness of the solution to (1.28), but because it also relies on luck in making a good guess, it is not generally available. In the examples below, the guess-and-verify method is often used. The reader should, however, be alerted to the fact that the objective functions and constraints of these problems have been especially rigged so that the method will work. Essentially there are only two classes of specifications of preferences and constraints for which the method will work, namely, variants of specifications with linear constraints and quadratic preferences or Cobb-Douglas constraints and logarithmic preferences.

In many problems, there is no unique way of defining states and controls, and several alternative definitions lead to the same solution of the problem. Sometimes the states and controls can be defined in such a way that x_t does not appear in the transition equation, so that $\partial g_t/\partial x_t \equiv 0$. In this case, the system (1.8a)–(1.8b) simplifies to

$$\frac{\partial r_t}{\partial u_t}(x_t, u_t) + \frac{\partial g_t}{\partial u_t}(u_t) \cdot \frac{\partial r_{t+1}(x_{t+1}, u_{t+1})}{\partial x_{t+1}} = 0, \qquad x_{t+1} = g_t(u_t).$$

The first equation is a version of what is called an Euler equation. Under circumstances in which the second equation can be inverted to yield u_t as a function of x_{t+1}, using the second equation to eliminate u_t from the first equation produces a second-order difference equation in x_t.

Most of the dynamic programming problems that we solve in this book are discounted dynamic programming problems.

1.4 Nonstochastic Examples

We now consider several examples of single-agent optimization problems that can be solved using dynamic programming.

Saving under Certainty

Consider the problem of a consumer in a nonrandom environment who seeks to maximize $\Sigma_{t=0}^{\infty}\beta^t u(c_t)$, $0 < \beta < 1$, subject to $A_{t+1} = R_t(A_t + y_t - c_t)$, A_0 given, where y_t, $t = 0, 1, \ldots$, is a known sequence of exponential order less than $1/\beta$ and R_t, $t = 0, 1, \ldots$, is a known and given sequence of one-period gross rates of return on nonlabor wealth. Here c_t is consumption, A_t is nonlabor wealth at the beginning of time t, and y_t is labor income at t. Labor income is assumed to be beyond the control of the agent. For concreteness let y_t equal λy_{t-1}, and say that $R_t = R > 0$ for all t, assuming that $R > \lambda > 0$. To rule out a strategy of infinite consumption supported by unbounded borrowing, we also impose the restriction that, for $t \geq 0$,

$$c_t + \sum_{j=1}^{\infty}\left(\prod_{k=0}^{j-1} R_{t+k}^{-1}\right) c_{t+j} = y_t + \sum_{j=1}^{\infty}\left(\prod_{k=0}^{j-1} R_{t+k}^{-1}\right) y_{t+j} + A_t. \tag{1.31}$$

We define the state of the system as (A_t, y_t, R_{t-1}) and define the control at t, u_t, as $R_t^{-1}A_{t+1} = A_t + y_t - c_t$. Evidently the control u_t is gross savings. The transition equation for A_t becomes $A_{t+1} = R_t u_t$, which does not involve the state at t. The function $r_t(x_t, u_t)$ becomes $\beta^t u(A_t + y_t - R_t^{-1}A_{t+1}) = \beta^t u(A_t + y_t - u_t)$. Bellman's equation becomes

$$v(A_t, y_t, R_{t-1}) = \max_{u_t}\{u(A_t + y_t - u_t) + \beta v(u_t R_t, y_{t+1}, R_t)\},$$

where $u_t = R_t^{-1}A_{t+1}$, $y_{t+1} = \lambda y_t$, $R_t = R$. Benveniste and Scheinkman's formula (1.30) gives $\partial v(A_t, y_t, R_{t-1})/\partial A_t = u'(c_t)$. The Euler equation for u_t then becomes

$$\begin{aligned}&-\beta^t u'(A_t + y_t - R_t^{-1}A_{t+1})\\ &+ \beta^{t+1}R_t u'(A_{t+1} + y_{t+1} - R_{t+1}^{-1}A_{t+2}) = 0\end{aligned}$$

or

$$-u'(c_t) + \beta R_t u'(c_{t+1}) = 0. \tag{1.32}$$

We seek a consumption plan that satisfies (1.32) and the "isoperimetric condition" (1.31).

As an example, suppose that $u(c_t) = \ln c_t$. Then (1.32) requires that

$$c_{t+j} = \beta^j\left(\prod_{k=0}^{j-1} R_{t+k}\right)c_t.$$

Substituting this into the left side of (1.31) gives $(1-\beta)^{-1}c_t$. Therefore (1.31) and (1.32) imply that

$$c_t = (1-\beta)\left[y_t + \sum_{j=1}^{\infty}\left(\prod_{k=0}^{j-1} R_{t+k}\right) y_{t+j} + A_t\right], \tag{1.33}$$

so that the agent always consumes a constant fraction of his or her total human and nonhuman wealth. Equation (1.33) is valid for any sequences $\{R_t\}_{t=0}^{\infty}$, $\{y_t\}_{t=0}^{\infty}$ such that the right side converges.

To specialize (1.33) to the case in which $y_t = \lambda y_{t-1}$ and $R_t = R$, write out (1.33) as

$$c_t = (1-\beta)(A_t + y_t + R_t^{-1}y_{t+1} + R_t^{-1}R_{t+1}^{-1}y_{t+2} + \ldots).$$

Repeatedly substituting $R_{t+1}^{-1} = R^{-1}$ and $y_{t+1} = \lambda y_t$ into the above equation gives

$$c_t = (1-\beta)(A_t + y_t + R^{-1}\lambda y_t + R^{-2}\lambda^2 y_t + \ldots)$$

or

$$c_t = (1-\beta)\left[A_t + y_t\left(\frac{1}{1-\lambda R^{-1}}\right)\right],$$

where we require that $\lambda R^{-1} < 1$. That is, income is assumed to grow at a rate less than the interest rate. In the decision rule stated above consumption varies directly with current income y_t, inversely with the currently observed interest rate R, and directly with the rate of growth of income λ.

Optimal Growth

A consumer aims to maximize

$$\sum_{t=0}^{\infty} \beta^t u(c_t), \qquad 0 < \beta < 1$$

$$\text{subject to} \quad c_t + k_{t+1} = f(k_t), \qquad k_0 > 0 \text{ given}, \qquad c_t \geq 0,$$

where $u'(0) = +\infty$, $u' > 0$, $u'' < 0$, $f'(0) = +\infty$, $f'(\infty) = 0$, $f' > 0$, and $f'' < 0$. Here c_t is consumption and k_t is the stock of capital. This is a version of the problem that was studied by T. C. Koopmans (1963) and David Cass (1965).

Let the state be defined as k_t and the control as k_{t+1}. Bellman's equation is then

$$v(k_t) = \max_{k_{t+1}}\{u[f(k_t) - k_{t+1}] + \beta v(k_{t+1})\}.$$

The first-order condition is

$$-u'[f(k_t) - k_{t+1}] + \beta v'(k_{t+1}) = 0. \tag{1.34}$$

Benveniste and Scheinkman's equation (1.30) implies that $v(k_t)$ is differentiable with

$$v'(k_t) = u'[f(k_t) - k_{t+1}]f'(k_t), \tag{1.35}$$

where k_{t+1} is evaluated at the optimum $k_{t+1} = h(k_t)$.

Because $u(\cdot)$ and $f(\cdot)$ are strictly concave, it follows that $v(k)$ is strictly concave. From this inference it follows that the optimal policy function, the solution $k_{t+1} = h(k_t)$ of (1.34), is a nondecreasing function of k_t.[2]

There is a maximum capital stock that can be sustained as a stationary equilibrium, namely that which would eventually emerge if c_t were to be zero for all t. If c_t were zero for all t, k_t would evolve according to the difference equation $k_{t+1} = f(k_t)$. Because $f'(0) = +\infty$, $f'' < 0$, and because $f'(\infty) = 0$, the equation $\bar{k} = f(\bar{k})$ has a unique positive solution. Evidently $k_{t+1} = f(k_t)$ converges to $\bar{k}$ as $t \to \infty$. [To verify this point, plot $f(k_t)$ against a 45° line.]

Let the system begin with $k_0 \in (0, \bar{k}]$. Then for $t \geq 1$, k_t must evidently remain in the bounded interval $[0, \bar{k}]$. Because the optimal policy function $h(k_t) = k_{t+1}$ is nondecreasing in k_t, it can be shown that $k_0, k_1, k_2, \ldots$, is a monotone, bounded sequence. On the one hand, suppose that $k_1 > k_0$. Then because $h(\cdot)$ is nondecreasing, we have $k_2 = h(k_1) \geq h(k_0) = k_1$, $k_3 = h(k_2) \geq h(k_1) = k_2$, and so on. On the other hand, suppose that $k_1 < k_0$. Then $k_2 = h(k_1) \leq h(k_0) = k_1$, $k_3 = h(k_2) \leq h(k_1) = k_2$, and so on. It follows that k_t is a monotone, bounded sequence. Inasmuch as monotone, bounded sequences converge, it follows that k_t converges to a limit point $k_\infty(k_0)$ as $t \to \infty$.

The preceding convergence argument leaves open the possibility that the limit point $k_\infty(k_0)$ depends on the starting point k_0. It does not do so, however, as the following argument verifies. Let k_∞ be a limit point. At the limit point, (1.34) and (1.35) hold, and $k_{t+1} = k_t = k_\infty$. The implication is that $\beta f'(k_\infty) = 1$, an equation that determines a unique optimal stationary value k_∞. Note that the "gross rate of return" $f'(k_\infty) = \beta^{-1}$ in the stationary state and is independent of the specifics of the current-period utility function and the production function. Note also that the optimal stationary capital stock depends on $f(\cdot)$ and β but not on $u(\cdot)$.

We now specialize this example by following Brock and Mirman (1972) and considering the particular functional forms $u(c) = \ln c$ and $f(k) = Ak^\alpha$, $A > 0, 0 < \alpha < 1$. We will use the guess-and-verify method for this problem. The guess may not seem an obvious one. The inspiration for the guess can be

2. From (1.35), we have that $v'(k)$ is continuous. This follows from the continuity of $h(k)$, $f(k)$, and $f'(k)$. For two levels k_i of k, $i = 1, 2$, consider the first-order condition $u'[f(k_i) - h(k_i)] = \beta V'[h(k_i)]$. Assume that $k_1 \geq k_2$ and that $h(k_1) < h(k_2)$. By strict concavity of $v(\cdot)$ and continuity of $v'(\cdot)$, it follows that (1) for all $h(k_i)$, $v'[h(k_i)]$ is well defined, and (2) $h(k_1) < h(k_2)$ implies $v'[h(k_1)] > v'[h(k_2)]$. Therefore, $u'[f(k_1) - h(k_1)] > u'[f(k_2) - h(k_2)]$. By strict concavity of u, the preceding inequality holds if and only if $f(k_1) - h(k_1) < f(k_2) - h(k_2)$, or equivalently, $0 < h(k_2) - h(k_1) < f(k_2) - f(k_1) \leq 0$. This is a contradiction produced by the assumption that for $k_1 \geq k_2$, $h(k_1) < h(k_2)$. Therefore $h(k)$ is nondecreasing in k. (The argument in this note was constructed by Rodolfo Manuelli.)

understood by working Exercise 1.1 at the end of the chapter. For this example we make the guess

(1.36) $$v(k) = E + F \ln k,$$

where E and F are undetermined coefficients. For this guess, the first-order necessary condition (1.34) implies the following formula for the optimal policy $\tilde{k} = h(k)$, where $\tilde{k}$ is next period's value and k is this period's value of the capital stock:

(1.37) $$\tilde{k} = \frac{\beta F}{1 + \beta F} A k^{\alpha}.$$

Substituting (1.37) into the right side of (1.35) gives

(1.38) $$v'(k) = (1 + \beta F)\alpha k^{-1}.$$

Differentiating (1.36) gives

(1.39) $$v'(k) = F k^{-1}.$$

Equating (1.38) and (1.39) permits one to solve for F, $F = \alpha/(1 - \alpha\beta)$. Substituting this expression for F back into (1.36) and (1.37) gives

(1.40) $$v(k) = E + \frac{\alpha}{1 - \alpha\beta} \ln k$$
$$\tilde{k} = A\beta\alpha k^{\alpha}.$$

The fact that expressions (1.38) and (1.39) for $v'(k)$ have identical functional forms both verifies the original guess (1.36) and permits one to solve for the undetermined coefficient F. An alternative procedure for verifying the guess involves substituting (1.37) into Bellman's functional equation and equating the result to the right side of (1.36). Solving the resulting equation for E and F again gives $F = \alpha/(1 - \alpha\beta)$ and now gives

$$E = (1 - \beta)^{-1}\left[\ln A(1 - \alpha\beta) + \frac{\beta\alpha}{1 - \alpha\beta} \ln A\beta\alpha\right].$$

In Exercise 1.1, the reader is asked to construct the same solution (1.37) to the functional equation, using the method of iterating on Bellman's equation (1.26) starting from $v_0(k) = 0$. For this purpose it is useful to note that the term $F = \alpha/(1 - \alpha\beta)$ that appears in (1.40) can be interpreted as a geometric sum $\alpha[1 + \alpha\beta + (\alpha\beta)^2 + \ldots]$.

Equation (1.40) shows that the optimal policy is to have capital move according to the difference equation $k_{t+1} = A\beta\alpha k_t^{\alpha}$, or $\ln k_{t+1} = \ln A\beta\alpha + \alpha \ln k_t$. Because $\alpha < 1$, we know that k_t converges as $t \to \infty$ for any positive initial value k_0. The stationary point is given by the solution of $k_\infty = A\beta\alpha k_\infty^{\alpha}$, or $k_\infty^{\alpha-1} = (A\beta\alpha)^{-1}$. Notice that this example obeys the general conclusion established above that k_∞ is determined from the solution of $\beta f'(k_\infty) = 1$.

1.5 The Optimal Linear Regulator Problem

We now consider a special class of dynamic programming problems in which the return functions r_t are quadratic and the transition functions g_t are linear. This specification leads to the widely used optimal linear regulator problem. We consider the special case in which the return functions r_t and transition functions g_t are both time invariant. The problem is to maximize over choice of $\{u_t\}_{t=0}^{\infty}$ the criterion

(1.41) $$\sum_{t=0}^{\infty} \{x_t' R x_t + u_t' Q u_t\},$$

subject to $x_{t+1} = Ax_t + Bu_t$, x_0 given. Here x_t is an $(n \times 1)$ vector of state variables, u_t is a $(k \times 1)$ vector of controls, R is a negative semidefinite symmetric matrix, Q is a negative definite symmetric matrix, A is an $(n \times n)$ matrix, and B is an $(n \times k)$ matrix. We guess that the value function is quadratic, $V(x) = x'Px$, where P is a negative semidefinite symmetric matrix.

Using the transition law to eliminate next period's state, Bellman's equation becomes

(1.42) $$x'Px = \max_u \{x'Rx + u'Qu + (Ax + Bu)'P(Ax + Bu)\}.$$

The first-order necessary condition for the maximum problem on the right side of (1.42) is

(1.43) $$(Q + B'PB)u = -B'PAx,$$

which implies the feedback rule for u:

(1.44) $$u = -(Q + B'PB)^{-1}B'PAx$$

or

(1.45) $$u = -Fx,$$

where $F = (Q + B'PB)^{-1}B'PA$. Substituting the optimizer (1.45) into the