Equation (4) admits a cycle of period 2 if there exist distinct numbers \( \xi_1 \) and \( \xi_2 \) such that \( f(\xi_1) = \xi_2 \) and \( f(\xi_2) = \xi_1 \). These numbers must be solutions of the equation \( x = f(f(x)) \). Since \( f(f(x)) \) is a polynomial of degree 4, it seems at first sight that we have to solve a rather difficult equation in order to find \( \xi_1 \) and \( \xi_2 \). Fortunately the equation simplifies because any solution of \( x = f(x) \) is also a solution of \( x = f(f(x)) \), so \( x - f(x) \) is a factor of the polynomial \( x - f(f(x)) \). A simple but tedious computation shows that

\[
\begin{align*}
g(x) &= a^2x^2 + a(b + 1)x + ac + b + 1
\end{align*}
\]

The cycle points are the roots of the equation \( g(x) = 0 \), which are

\[
\begin{align*}
\xi_{1,2} &= \frac{-(b + 1) \pm \sqrt{(b - 1)^2 - 4ac - 4}}{2a}
\end{align*}
\]

These roots exist and are distinct if and only if \((b - 1)^2 > 4ac + 4\). Hence, if there is a cycle of period 2, the equilibrium points \( x_1 \) and \( x_2 \) also exist, and are both unstable. (See also Problem 1.)

Because \( f'(x) = 2ax + b \), while \( \xi_1 + \xi_2 = -(b + 1)/a \) and \( \xi_1\xi_2 = (ac + b + 1)/a^2 \), a simple calculation shows that \( f'(\xi_1)f'(\xi_2) = 4ac - (b - 1)^2 + 5 \). Then

\[
\begin{align*}
f'(\xi_1)f'(\xi_2) < 1 \iff 4 - (b - 1)^2 - 4ac < 0
\end{align*}
\]

It follows that if both inequalities on the right are satisfied, then equation (4) admits a stable cycle of period 2. (The first inequality on the right is precisely the necessary and sufficient condition for a period 2 cycle to exist.)

PROBLEMS FOR SECTION 11.7

1. Show that if \( f : I \to I \) is continuous and the difference equation \( x_{n+1} = f(x_n) \) admits a cycle \( \xi_1, \xi_2 \) of period 2, it also has at least one equilibrium solution between \( \xi_1 \) and \( \xi_2 \). (Hint: Consider the function \( f(x) - x \) over the interval with endpoints \( \xi_1 \) and \( \xi_2 \).)

2. A solution \( x^* \) of the equation \( x = f(x) \) can be viewed as an equilibrium solution of the difference equation

\[
\begin{align*}
x_{n+1} &= f(x_n) \quad (n)
\end{align*}
\]

If this equilibrium is stable and \( x_0 \) is a sufficiently good approximation to \( x^* \), then the solution \( x_0, x_1, x_2, \ldots \) of (n) starting from \( x_0 \) will converge to \( x^* \).

(a) Use this technique to determine the negative solution of \( x = e^x - 3 \) to at least three decimal places.

(b) The equation \( x = e^x - 3 \) also has a positive solution, but this is an unstable equilibrium of \( x_{n+1} = e^{x_n} - 3 \). Explain how nevertheless we can find the positive solution by rewriting the equation and using the same technique as above.

3. The function \( f \) in Fig. 4 is given by \( f(x) = -x^2 + ax - 4x - 4/5 \). Find the values of the cycle points \( \xi_1 \) and \( \xi_2 \), and use (5) to determine whether the cycle is stable. It is clear from the figure that the difference equation \( x_{n+1} = f(x_n) \) has two equilibrium states. Find these equilibria, show that they are both unstable, and verify the result in Problem 1.

DISCRETE TIME OPTIMIZATION

In science, what is susceptible to proof
must not be believed without proof.\(^1\)
—R. Dedekind (1887)

This chapter gives a brief introduction to discrete time dynamic optimization problems. The term dynamic is used because the problems involve systems evolving over time. Time is here 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 dynamic systems whose development is governed by difference equations.

If the horizon is finite, such dynamic 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 11.1 on dynamic programming studies a standard problem with one state and one control variable.

In the economics literature a dynamic programming version of the Euler equation in continuous time control theory is much used. Section 11.2 gives a brief description.

When discussing optimization problems in discrete time, economists often prefer models with an infinite time horizon, just as they do in continuous time. Section 12.3 treats such models. The fundamental result is the Bellman equation.

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. Sections 12.4 and 12.5 set out the relevant discrete time maximum principle, first for a single state variable, then for many. In contrast to the continuous time maximum principle, the Hamiltonian is not necessarily maximized at the optimal control. Section 12.5 also presents a very brief discussion of infinite horizon problems in this setting.

Section 12.6 offers an introduction to stochastic dynamic programming, including the stochastic Euler equation that plays such a prominent role in current macroeconomic theory. The concluding Section 12.7 is devoted to the important case of stationary problems with an infinite horizon. (Sections 12.6 and 12.7 are the only parts of the book that rely on some knowledge of probability theory, though only at a basic level.)

\(^1\) There is no ideal English translation of the German original: "Was beweisbar ist, soll in der Wissenschaft nicht ohne Beweis geäußert werden."
12.1 Dynamic Programming

Consider a system that changes at discrete times $t = 0, 1, \ldots, T$. Suppose the state of the system at time $t$ is characterized by a real number $x_t$. For example, $x_t$ might be the quantity of grain that is stockpiled at time $t$. 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 $u_t$, which can be chosen freely from a given set $U$, called the control region. For example, $u_t$ might be the quantity of grain removed from the store $s_t$ at time $t$. The controls influence the evolution of the system through a difference equation

$$x_{t+1} = g(t, x_t, u_t), \quad x_0 \text{ given}, \quad u_t \in U \tag{1}$$

where $g$ is a given function. Thus, we assume that the state of the system at time $t + 1$ depends explicitly 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 $t$.

Suppose that we choose values for $u_0, u_1, \ldots, u_{T-1}$. Then (1) gives $x_1 = g(0, x_0, u_0)$, $x_2 = g(1, x_1, u_1)$, then next $x_3 = g(2, x_2, u_2)$, etc. 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, $u_0, \ldots, u_{T-1}$. Each choice of $(u_0, u_1, \ldots, u_{T-1})$ gives rise to a sequence $(x_0, x_1, \ldots, x_T)$, for instance path 1 in Fig. 1. A different choice of $(u_0, u_1, \ldots, u_{T-1})$ would give another path, such as path 2 in the figure. Such controls $u_t$ that depend only on time are often called open-loop controls.

![Figure 1](https://example.com/figure1.png)

**Figure 1** Different evolutions of system (1)

Different paths will usually have different utility or value. Assume that there is a function $f(t, x, u)$ of three variables such that the utility associated with a given path is represented by the sum

$$\sum_{t=0}^{T} f(t, x_t, u_t) \tag{2}$$

The sum is called the **objective function**, and it represents the sum of utilities (values) obtained at each point of time.

**NOTE 1** 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 of (2) in which $f(T, x_T, u_T) = S(x_T)$. ($S$ is often called a scrap value function.)

Suppose that we choose values for $u_0, u_1, \ldots, u_{T-1}, u_T$, all from the set $U$, as specified in (1). The initial state $x_0$ is given, and as explained above, (1) gives $x_1, \ldots, x_T$. Let us denote corresponding pairs $(x_0, x_1, \ldots, x_T)$, $(u_0, u_1, \ldots, u_T)$ by $(x_0^*, u_0^*)$, and call them **admissible sequence pairs**. For each admissible sequence pair the objective function has a definite value. We shall study the following problem:

**Among all admissible sequence pairs $(x_0^*, u_0^*)$ find one, $(x_0^*, u_0^*)$, that makes the value of the objective function as large as possible.**

Such an admissible sequence pair is called an **optimal pair**, and the corresponding control sequence $(u_0^*, u_1^*, \ldots, u_T^*)$ is called an **optimal control**. The discrete time optimization problem can be briefly formulated as

$$\max \sum_{t=0}^{T} f(t, x_t, u_t) \text{ subject to } x_{t+1} = g(t, x_t, u_t), \quad x_0 \text{ given}, \quad u_t \in U \tag{2}$$

**EXAMPLE 1** Let $s_t$ be an individual’s wealth at time $t$. At each point of time $t$, the individual has to decide the proportion $u_t$ of $s_t$ to consume, leaving the remaining proportion $1 - u_t$ for savings. Assume that wealth earns interest at the rate $r > 0$. After $u_t s_t$ has been withdrawn for consumption, the remaining stock of wealth is $(1 - u_t) s_t$. Because of interest, this grows to the amount $x_{t+1} = \rho (1 - u_t)s_t$ at the beginning of period $t + 1$. This equation holds for $t = 0, \ldots, T - 1$, with $x_0$ a positive constant. Suppose that the utility of consuming $c_t = u_t s_t$ is $U(t, c_t)$. Then the total utility over periods $t = 0, \ldots, T$ is

$$\sum_{t=0}^{T} U(t, u_t s_t)$$

The problem facing the individual is therefore the following:

$$\max \sum_{t=0}^{T} U(t, u_t s_t) \text{ subject to } x_{t+1} = \rho (1 - u_t)s_t, \quad t = 0, \ldots, T - 1 \tag{3}$$

with $x_0$ given and with $u_t$ in $[0, 1]$ for $t = 0, \ldots, T$. This is a standard dynamic optimization problem of the type described above. (See Problems 2, 3, and 6.)

### The Value Function and its Properties

Returning to the general problem described by (2), suppose that at time $t$ $x$ the state of the system is $x$ (any given real number). The best we can do in the remaining periods is to choose $u_t, u_{t+1}, \ldots, u_T$ (and thereby also $x_{t+1}, \ldots, x_T$) to maximize $\sum_{t=0}^{T} f(t, x_t, u_t)$ with $x_T = x$. We define the **(optimal) value function** for the problem at time $t$ by

$$J_t(x) = \max_{u_t, \ldots, u_T \in U} \sum_{t=0}^{T} f(t, x_t, u_t) \tag{3}$$

where

$$x_T = x \quad \text{and} \quad x_{t+1} = g(t, x_t, u_t) \quad \text{for} \ t > s, \quad u_t \in U \tag{4}$$

At the terminal time $t = T$, definition (3) implies that $J_T(x) = \max_{u_T \in U} f(T, x, u_T)$.

1 We assume that the maximum in (3) is attained. This is true if, for example, the functions $f$ and $g$ are continuous and $U$ is compact.
We now prove an important property of the value function. Suppose that at time \( t = s \) \((s < T)\) we are in state \( x_s = x \). What is the optimal choice for \( u_s \)? If we choose \( u_s = u \), then at time \( t = s \) we obtain the immediate reward \( f(x_s, u_s) \), and according to (4), the state at time \( s + 1 \) will be \( x_{s+1} = g(x_s, u) \). Using definition (3) again, the highest obtainable value of the total reward \( \sum_{t=s+1}^{T} f(x_t, u_t) \) from time \( s + 1 \) to time \( T \), starting from the state \( x_{s+1} = g(x_s, u) \). Hence, the best choice of \( u \) is \( u_s = u \) at time \( s \) must be a value of \( u \) that maximizes the sum

\[
f(x_s, u_s) + J_{s+1}(g(x_s, u))
\]

This leads to the following general result:

**THEOREM 12.1.1 (FUNDAMENTAL EQUATIONS OF DYNAMIC PROGRAMMING)**

For each \( s = 0, 1, \ldots, T - 1 \), let \( J_s(x) \) be the value function (3) for the problem

\[
\max \sum_{t=0}^{T} f(x_t, u_t) \quad \text{subject to} \quad x_{t+1} = g(x_t, u_t), \quad u_t \in U
\]

with \( x_0 \) given. Then the sequence of value functions satisfies the equations

\[
J_s(x) = \max_{u \in U} \left[ f(x, u) + J_{s+1}(g(x, u)) \right], \quad s = 0, 1, \ldots, T - 1
\]

\[
J_T(x) = \max_{u \in U} f(T, x, u)
\]

**NOTE 2** If we minimize rather than maximize the sum in (5), then Theorem 12.1.1 holds with "max" replaced by "min" in (3), (6) and (7). This is because minimizing \( f \) is equivalent to maximizing \( -f \).

**NOTE 3** Let \( \mathcal{X}_s(x_0) \) 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 \( J_s(x) \) for \( x \in \mathcal{X}_s(x_0) \) are relevant.

Theorem 12.1.1 is the basic tool for solving dynamic optimization problems. It is used as follows: First find the function \( J_T(x) \) by (7). The maximizing value of \( u \) depends (usually) on \( x \), and is denoted by \( u^*_T(x) \). The next step is to use (6) to determine \( J_{T-1}(x) \) and the corresponding maximizing control \( u^*_{T-1}(x) \). Then work backwards in this fashion to determine recursively all the value functions \( J_T(x), \ldots, J_0(x) \) and the maximizers \( u^*_T(x), \ldots, u^*_0(x) \). This allows us to construct the solution of the original optimization problem.

Since the state at \( t = 0 \) is \( x_0 \), the 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 \( x_1^* = g(x_0, u^*_0(x_0)) \). Then \( u^*_1(x_1^*) \) is the best choice of \( u_1 \), and this choice determines \( x_2^* \) by (1). Then again, \( u^*_2(x_2^*) \) is the best choice of \( u_2 \), and so on.

### EXAMPLE 2

Use Theorem 12.1.1 to solve the problem

\[
\max \sum_{t=0}^{3} (1 + x_t - u_t^2), \quad x_{t+1} = x_t + u_t, \quad t = 0, 1, 2, \quad x_0 = 0, \quad u_t \in \mathbb{R}
\]

**Solution:** Here \( T = 3 \), \( f(t, x, u) = 1 + x - u^2 \), and \( g(t, x, u) = x + u \). Consider first (7) 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 at \( u = 0 \). Hence, in the notation introduced above,

\[
J_3(x) = 1 + x, \quad \text{with} \quad u^*_3(x) = 0
\]

For \( s = 2 \), the function to be maximized in (6) is \( h_2(u) = 1 + x - u^2 + J_3(x + u) \), where (5) implies that \( J_3(x + u) = 1 + (x + u) \). Thus, \( h_2(u) = 1 + x - u^2 + 1 + (x + u) = 2 + 2x + u - u^2 \). The function \( h_2 \) is concave in \( u \), and \( h_2'(u) = 2 + 2x = 0 \) for \( u = -x/2 \), so this is the optimal choice of \( u \). Then the maximum value of \( h_2(x) \) is \( 2 + 2x + 1/2 - 1/4 \). Hence,

\[
J_2(x) = \frac{1}{4} + 2x, \quad \text{with} \quad u^*_2(x) = -\frac{x}{2}
\]

For \( s = 1 \), the function to be maximized in (6) is given by \( h_1(u) = 1 + x - u^2 + J_2(x + u) = 1 + x - u^2 + 9/4 + 2(x + u) = 13/4 + 3x + 2u - u^2 \). Because \( h_1 \) is concave and \( h_1'(u) = 2 - 2u = 0 \) for \( u = 1 \), the maximum value of \( h_1(u) \) is \( 13/4 + 3x + 2 - 1 = 17/4 + 3x \), so

\[
J_1(x) = \frac{17}{4} + 3x, \quad \text{with} \quad u^*_1(x) = 1
\]

Finally, for \( s = 0 \), the function to be maximized is \( h_0(u) = 1 + x - u^2 + J_1(x + u) = 1 + x - u^2 + 17/4 + 3(x + u) = 21/4 + 4x + 3u - u^2 \). The function \( h_0 \) is concave and \( h_0'(u) = 3 - 2u = 0 \) for \( u = 3/2 \), so the maximum value of \( h_0(u) \) is \( 21/4 + 4(3/2) = 15/2 + 4x \). Thus,

\[
J_0(x) = \frac{15}{4} + 4x, \quad \text{with} \quad u^*_0(x) = \frac{3}{2}
\]

In this particular case the optimal choices of the controls are constants, independent of the states. 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 \( 15/2 \).

**Alternative solution:** In simple cases like this, a dynamic optimization problem can be solved quite easily by ordinary calculus methods. By letting \( r = 0, 1 \), and 2 in the difference equation \( x_{t+1} = x_t + u_t \), we get \( x_1 = x_0 + u_0 = u_0, \ x_2 = x_1 + u_1 = u_0 + u_1, \ x_3 = x_2 + u_2 = u_0 + u_1 + u_2 \). Using these results, the objective function becomes the following function of \( u_0, u_1, u_2, \) and \( x_0 \):

\[
I = (1 - u_0^2) + (1 + u_0 + u_1 - u_0^2) + (1 + u_0 + u_1 + u_2 - u_0^2)
\]

\[
= 4 + 3u_0 - u_0^2 + 2u_1 - u_1^2 + u_2 - u_2^2
\]


The problem has been reduced to that of maximizing $J$ with respect to the control variables $u_0, u_1, u_2,$ and $u_3$. We see that $J$ is a sum of concave functions and so is concave. Hence, a stationary point will maximize $J$. The first-order derivatives of $I$ are

$$\frac{\partial I}{\partial u_0} = 3 - 2u_0, \quad \frac{\partial I}{\partial u_1} = 2 - 2u_1, \quad \frac{\partial I}{\partial u_2} = 1 - 2u_2, \quad \frac{\partial I}{\partial u_3} = -2u_3$$

Equating these partial derivatives to zero yields the unique stationary point $(u_0, u_1, u_2, u_3) = (\frac{3}{2}, 1, \frac{1}{2}, 0)$. This gives the same solution as the one obtained by using Theorem 12.1.1.

In principle, all deterministic finite horizon dynamic problems can be solved in this alternative way using ordinary calculus. But the method becomes very unwieldy if the horizon $T$ is large, or if there is a stochastic optimization problem of the kind considered in Sections 12.6–12.7 below.

In the next example the terminal time is an arbitrarily given natural number and the optimal control turns out to depend on the state of the system.

**EXAMPLE 3** Solve the following problem:

$$\max_{x_0 \geq 0} \sum_{t=0}^{T-1} \left( -\frac{1}{2}ux_t + \ln x_t \right), \quad x_{t+1} = x_t(1 + u_t x_t), \quad x_0 \text{ positive constant}, \quad u_t \geq 0$$

**Solution:** Because $x_0 > 0$ and $u_t \geq 0$, we have $x_t > 0$ for all $t$. Now $f_0(T, x, u) = \ln x$ is independent of $u$, so $J_T(x) = \ln x$, and any $u_T$ is optimal.

Next, putting $s = T - 1$ in (6) yields

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

The maximum of the concave function $h(u) = -\frac{1}{2}ux + \ln x + \ln(1 + ux)$ is at the point where its derivative is 0. This gives $h'(u) = -\frac{1}{2}x + x/(1 + ux) = 0$, or (since $x > 0$), $u = 1/2x$. Then $h(1/2x) = \ln x - 1/3 + \ln(3/2)$. Hence,

$$J_{T-1}(x) = h(1/2x) = \ln x + C, \quad x = -1/3 + \ln(3/2), \quad \text{and } u_{T-1}^*(x) = 1/2x$$

The next step is to use (6) for $s = T - 2$:

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

Again $u = u_{T-2}^*(x) = 1/2x$ gives the maximum because the first-order condition is the same, and we get

$$J_{T-2}(x) = \ln x + 2C, \quad x = -1/3 + \ln(3/2), \quad \text{and } u_{T-2}^*(x) = 1/2x$$

This pattern continues and so, for $k = 0, 1, \ldots, T$ we get

$$J_{T-k}(x) = \ln x + kC, \quad C = -1/3 + \ln(3/2), \quad \text{and } u_{T-k}^*(x) = 1/2x$$

So far we have been working backwards from time $T$ to time 0. Putting $t = T - k$ for each $k$, we find that $J_k(x) = \ln x + (T - t)C$ and $u_k^* = 1/2x$ for $t = 0, 1, \ldots, T$.

Finally, inserting $u_T^* = 1/2x$ in the difference equation gives $x_{T+1}^* = \left(\frac{3}{2}\right)x_T^*$. So $x_T^* = \left(\frac{3}{2}\right)^T x_0$, with $u_t = \left(\frac{3}{2}\right)^T / 2x_0$ as optimal control values.

**NOTE 4** Theorem 12.1.1 also holds if the control region is not a fixed set $U$, but instead a set $U(t, x)$ that depends on $(t, x)$. Then the maximization in (2), (3), and (5) is carried out for $u_t$ in $U(t, x)$. In (6) and (7), the maximization is carried out for $u \in U(x, x)$ and $u \in U(T, x)$, respectively. Frequently, the set $U(t, x)$ is determined by one or more inequalities of the form $h(t, x, u) \leq 0$, for some function $h$ that is continuous in $(x, u)$. If $U(t, x)$ is empty, then by convention, the maximum over $U(t, x)$ is set equal to $-\infty$.

**NOTE 5** In the above formulation, the state $x$ and the control $u$ may well be vectors, in say $\mathbb{R}^n$ and $\mathbb{R}^m$, 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 12.1.1 (except that we would use boldface letters for $x, u$, and $g$).

**EXAMPLE 4** Let $x_t$ denote the value of an investor’s assets at time $t$, and let $u_t$ be consumption. Suppose that assets at time $t + 1$ are proportional to savings $s_t - u_t$ at $t$, with a factor of proportionality depending on $t$, i.e.

$$x_{t+1} = a_t(s_t - u_t), \quad a_t \text{ given positive numbers}$$

Assume that the initial assets, $x_0$, are positive. The utility associated with a level of consumption $u$ during one period is supposed to be $u^{1-r}$, while the utility of the assets at time $T$ is $A x_T^{1-r}$. Here $A$ is a positive constant and $r \in (0, 1)$. The investor wants to maximize the discounted value of the sum of utility from consumption and terminal assets. Define $\beta = 1/(1 + r)$, where $r$ is the rate of discount. Assume that both savings and consumption must be positive each period, so $0 < u_t < x_t$. The investor’s problem is thus:

$$\max_{\{u_t\}_{t=0}^{T-1}} \sum_{t=0}^{T-1} \beta^t u_t^{1-r} + \beta^T A x_T^{1-r}, \quad x_{T+1} = a_T(s_T - u_T), \quad u_t \in (0, x_t)$$

**Solution:** We apply Theorem 12.1.1, as amended by Note 4, with the control region $U(t, x)$ given by the open interval $(0, x)$. So $f(t, x, u) = \beta^t u^{1-r}$ for $t = 0, 1, \ldots, T - 1$, whereas $f(T, x, u) = \beta^T A x^{1-r}$. Since this function does not depend on $u$, (7) yields

$$J_T(x) = \max_{x \in (0, x)} \beta^T A x^{1-r} = \beta^T A x^{1-r}$$

$$J_{T-1}(x) = \max_{x \in (0, x)} \beta^T A x^{1-r} = \beta^T A x^{1-r}$$
and any \( u_T \) in \((0, x)\) is optimal. Moreover, equation (6) yields
\[
J_T(x) = \max_{u \in (0, x)} \left[ \beta^u u^{-1 - \gamma} + J_{T + 1}(a_T(x - u)) \right]
\]
\[
\text{(iii)}
\]
In particular, (ii) gives \( J_T(\alpha T - 1(x - u)) = \beta T \alpha T - 1(x - u)^{-1 - \gamma} \), so
\[
J_{T - 1}(x) = \beta^{T - 1} \max_{u \in (0, x)} \left[ u^{-1 - \gamma} + \beta C_{T - 1}(x - u)^{-1 - \gamma} \right]
\]
\[
\text{(iv)}
\]
Let \( h(u) = u^{-1 - \gamma} + c'(x - u)^{-1 - \gamma} \) denote the maximand, as a function of \( u \) in \((0, x)\), where \( c' = \beta C_{T - 1}^{1/\gamma} \). Then \( h'(u) = (1 - \gamma)u^{-1 - \gamma} - (1 - \gamma)c'(x - u)^{-\gamma} = 0 \) when \( u = c'(x - u)^{-\gamma} \) and so \( u = (x - u)/c' \), implying that
\[
u_{T - 1} = u = x/u, \quad \text{where} \quad u = 1 + c' = 1 + (\beta C_{T - 1}^{1/\gamma})^{-1} = C_{T - 1}^{1/\gamma}
\]
\[
\text{(v)}
\]
for a suitably defined constant \( C_{T - 1} \). Because \( \gamma \in (0, 1) \) and \( c' > 0 \), the function \( h \) is easily seen to be concave over \((0, x)\). So the value of \( u \) given in \( (v) \) does maximize \( h(u) \), Then, because \( \beta C_{T - 1}^{1/\gamma} = c' = (w - 1)/\gamma \), choosing the value \( x/u \) of \( u_{T - 1} \) gives
\[
h(x/u) = x^{1 - \gamma}w^{-1} + (w - 1)/\gamma(x(1 - w)^{-1})^{-1 - \gamma}
= x^{1 - \gamma}w^{-1} + (w - 1)/\gamma(w - 1)^{-1 - \gamma} = x^{1 - \gamma}w^{\gamma} = x^{1 - \gamma}C_{T - 1}
\]
Hence, by \( (iv) \),
\[
J_{T - 1}(x) = \beta^{T - 1} C_{T - 1} x^{1 - \gamma}
\]
\[
\text{(vi)}
\]
Notice that \( J_{T - 1}(x) \) has the same form as \( J_T(x) \). Next, substitute \( s = T - 2 \) in \( (iii) \) to get:
\[
J_{T - 2}(x) = \beta^{T - 2} \max_{u \in (0, x)} \left[ u^{-1 - \gamma} + \beta C_{T - 1}^{1/\gamma} C_{T - 2} (x - u)^{-1 - \gamma} \right]
\]
Comparing this with \( (iv) \), from \( (v) \) we see that the maximum value is attained for
\[
u_{T - 2} = u = x/C_{T - 2}^{1/\gamma}, \quad \text{where} \quad C_{T - 2}^{1/\gamma} = 1 + (\beta C_{T - 1}^{1/\gamma})^{-1}
\]
and that \( J_{T - 1}(x) = \beta^{T - 2} C_{T - 1}^{1/\gamma} x^{1 - \gamma} \).

We can obviously go backwards repeatedly in this way and, for every \( t \), obtain
\[
J_t(x) = \beta^t C_t x^{1 - \gamma}
\]
\[
\text{(vii)}
\]
From \( (ii) \), \( C_T = A \), while \( C_t \) for \( t < T \) is determined by backward recursion using the first-order difference equation
\[
C_t^{1/\gamma} + (\beta C_{t + 1}^{1/\gamma})^{1/\gamma} = A^{1/\gamma} C_{t + 1}^{1/\gamma}
\]
\[
\text{(viii)}
\]
that is linear in \( C_t^{1/\gamma} \). The optimal control is
\[
\hat{u}_t(x) = x/C_t^{1/\gamma}, \quad t < T
\]
\[
\text{(x)}
\]
We find the optimal path by successively inserting \( \hat{u}_0, \hat{u}_1, \ldots \) into the difference equation \( (i) \) for \( x_t \).

We can obtain an explicit solution in the special case when \( \alpha = a \) for all \( t \). Then \( (vii) \) reduces to
\[
C_t^{1/\gamma} = C_{T - 1}^{1/\gamma} = 1
\]
\[
\text{where} \quad \alpha = (\beta a^{1 - \gamma})^{1/\gamma}
\]
\[
\text{(xi)}
\]
This is a first-order linear difference equation with constant coefficients. Using \( C_T = A \), and solving the equation for \( C_t^{1/\gamma} \), we obtain
\[
C_t^{1/\gamma} = A^{1/\gamma} C_{T - 1}^{1/\gamma} + \left. \frac{1 - \alpha^{t - 1}}{1 - \alpha} \right|_{t = T}, \quad T - 1, \ldots, 0
\]
\[
\text{(xii)}
\]
\[\text{NOTE 6}\]

Controls \( u_t(x_t) \) that depend on the state \( x_t \) of the system are called closed-loop controls, whereas controls \( u_t \) that only depend on time are called open-loop controls.

Excepts in rare special cases, the controls \( u_0', \ldots, u_{T - 1}' \) that yield the maximum value \( J_T(x) \) in \( (iii) \) evidently do depend on \( x_T \). In particular the first control \( u_0' \) does so, i.e. \( u_0' = u_0(x_0) \). So, determining the functions \( J_t(x) \) defined in \( (iii) \) requires finding optimal closed-loop controls \( u_t'(x_t) \), \( t = 0, 1, \ldots, T \).

Given the initial state \( x_0 \) and any sequence of closed-loop controls \( u_t'(x_t) \), the evolution of the state \( x_t \) is uniquely determined by the difference equation
\[
x_{t + 1} = g(t, x_t, u_t(x_t)) \quad \text{for} \quad t = 0, 1, \ldots, T\text{.}
\]
\[
\text{(xiii)}
\]
Let us denote by \( \hat{u}_t = u_t(x_t) \) the control values (numbers) generated by this particular sequence of states \( x_t \). Next, insert these numbers \( \hat{u}_t \) into the difference equation:
\[
x_{t + 1} = g(t, x, \hat{u}_t) \quad \text{for} \quad t = 0, 1, \ldots, T\text{.}
\]
\[
\text{(xiv)}
\]
This obviously has exactly the same solution as equation \( (x) \).

Hence, we get the same result whether we insert the closed-loop controls \( u_t'(x_t) \) or the equivalent open-loop controls \( u_t \). 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 \( t \), there is an unexpected disturbance to the state \( x_t \) obtained from the difference equation, which has the effect of changing the state to \( x_t' \). Then \( u_t'(x_t') \) still gives the optimal control to be used at that time, provided we assume that no further disturbances will occur.

\textbf{PROBLEMS FOR SECTION 12.1}

\( \text{Problem 1.} \) Use Theorem 12.1.1 to solve the problem
\[
\max_{\hat{u}, u} \sum_{t = 0}^{T - 1} [1 - (x_t^2 + 2u_t^2)] \quad x_{t + 1} = x_t - u_t, \quad t = 0, 1, \ldots \text{.}
\]
\[
\text{where} \quad x_0 = 5 \quad \text{and} \quad u_t \in \mathbb{R} \text{. (Compute } J_t(x) \text{ and } u_t'(x) \text{ for } s = 2, 1, 0, \ldots \text{.)}
\]
\[
\text{Use the difference equation in } (x) \text{ to compute } x_1 \text{ and } x_2 \text{ in terms of } u_0 \text{ and } u_1 \text{ (with } x_0 = 5 \text{),}
\]
\[
\text{and find the sum in } (x) \text{ as a function } S \text{ of } u_0, u_1, \text{ and } u_2 \text{. Next, maximize this function as in Example 2.}
\]
2. Consider the problem
\[
\max_{x_0 \in [0,1]} \sum_{t=0}^{T} \left( \frac{1}{1 + r} \right)^t \sqrt{n_t x_t}, \quad x_{t+1} = \rho (1 - u_t) x_t, \quad t = 0, \ldots, T - 1, \quad x_0 > 0
\]
where \(r\) is the rate of discount. Compute \(I_c(x)\) and \(u^*_c(x)\) for \(s = T, T - 1, T - 2\).

3. (a) Replace the utility function in Problem 2 by \(\sum_{t=0}^{T} (1 + r)^{-t} u_t x_t\). Compute \(J_c(x), u_t^c(x)\), \(J_{c-1}(x)\), and \(u^c_{t-1}(x)\) for \(x \geq 0\).
(b) Prove that there exist constants \(P_c\) (depending on \(\rho\) and \(r\)) such that \(J_c(x) = P_c x\) for \(s = 0, 1, \ldots, T\). Find \(J_0(x)\) and optimal values of \(u_0, u_1, \ldots, u_T\).

4. Consider the problem
\[
\max_{x_0 \in (0,1]} \int_0^T 3 - u_t x_t dt, \quad x_{t+1} = u_t x_t, \quad t = 0, \ldots, T - 1, \quad x_0 \geq 0
\]
(a) Compute the value functions \(J_c(x), J_{c-1}(x), J_{c-2}(x), \) and the corresponding control functions, \(u^c_t(x), u^c_{t-1}(x), \) and \(u^c_{t-2}(x)\).
(b) Find an expression for \(J_{c-1}(x)\) for \(n = 0, 1, \ldots, T\), and the corresponding optimal controls.

5. Solve the problem
\[
\max_{x_0 \in (0,1]} \int_0^T \left( \frac{1}{2} u_t^2 + \ln x_t \right) dt, \quad x_{t+1} = x_t (1 + u_t), \quad t = 0, \ldots, T - 1, \quad x_0 > 0 \text{ given}
\]
(a) Write down the fundamental equations for the problem
\[
\max_{u_t \in [0,1]} \int_0^T \left( \frac{1}{2} u_t^2 - u_t \right) dt, \quad x_{t+1} = 2x_t (1 + u_t), \quad t = 0, 1, \ldots, T - 1, \quad x_0 = 0
\]
(b) Prove that the value function for the problem is given by
\[
J_{c-1}(x) = (2^{n+1} - 1) x + \sum_{j=0}^{n} (2^j - 1) x^j, \quad n = 0, 1, \ldots, T
\]
Determine the optimal controls \(u_t = u^c_t\) and the maximum value \(V = J_0(0)\).

6. (a) Consider the problem
\[
\max_{x_0 \in (0,1]} \int_0^T \left( e^{-\alpha t} - \alpha e^{-\gamma t} \right) dt, \quad x_{t+1} = 2x_t - u_t, \quad t = 0, 1, \ldots, T - 1, \quad x_0 \geq 0
\]
where \(\alpha\) and \(\gamma\) are positive constants. Compute \(J_c(x), J_{c-1}(x),\) and \(J_{c-2}(x)\).
(b) Prove that \(J_c(x)\) can be written in the form \(J_c(x) = -\alpha x e^{-\gamma t}\), and find a difference equation for \(u_t\).

8. Consider the following special case of Problem 2, where \(r = 0\):
\[
\max_{x_0 \in [0,1]} \sum_{t=0}^{T} \sqrt{n_t x_t}, \quad x_{t+1} = \rho (1 - u_t) x_t, \quad t = 0, \ldots, T - 1, \quad x_0 > 0
\]
(a) Compute \(J_0(x), J_{c-1}(x), J_{c-2}(x)\). (Hint: Prove that \(\max_{u_t \in [0,1]} \left[ \sqrt{u} + \sqrt{1 - u} \right] = \sqrt{1 + \lambda^2}\) with \(\lambda = 1/(1 + \lambda^2)\).)
(b) Show that the optimal control function is \(u_t(x) = 1/(1 + \rho^2 + \cdots + \rho^{T-t})\), and find the corresponding \(J_0(x), s = 1, 2, \ldots, T\).

### 12.2 The Euler Equation

The economics literature sometimes considers the following formulation of the basic dynamic programming problem without an explicit control variable (e.g., Stokey et al. (1989))
\[
\max_{x_0 \in (0,1]} \sum_{t=0}^{T} F(t, x_t, x_{t+1}), \quad x_0 \text{ given and } x_1, x_2, \ldots, x_T, x_{T+1} \text{ vary freely in } \mathbb{R}
\]
(1)
Here 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 (1) becomes a standard dynamic programming problem with \(U = \mathbb{R}\). On the other hand, the dynamic optimization problem (12.1.2) can usually be formulated as a problem of the type (1). Suppose, in particular, that 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 \(U\), which we denote by \(u_t = \psi(t, x_t, x_{t+1})\). Now define the function \(F\) by \(F(t, x_t, x_{t+1} = f(t, x_t, \psi(t, x_t, x_{t+1}))\) for \(t < T\), and \(F(T, x_T, x_{T+1}) = \max_{x_{T+1}} f(T, x_T, u)\). Then problem (12.1.2) becomes precisely the same as problem (1).

If there is more than one value of \(u\) in \(U\) such that \(g(t, x_t, u) = x_{t+1}\), let \(u_t\) be a value of \(u\) that maximizes \(f(t, x_t, u)\), i.e. choose the best \(u\) that leads from \(x_t\) to \(x_{t+1}\). Then, for each \(t = 0, 1, \ldots, T\), we have
\[
F(t, x_t, x_{t+1}) = \max \{ f(t, x_t, u) : x_{t+1} = g(t, x_t, u), \ u \in U \}
\]
(2)
Let \((x^*_0, \ldots, x^*_T)\) be an optimal solution of problem (1). Then \((x^*_0, \ldots, x^*_T)\) is a maximum point for the objective function \(S(x_0, \ldots, x_{T+1}) = \sum_{t=0}^{T} F(t, x_t, x_{t+1})\), and by the usual first-order condition we must have \(S_t(x^*_0, \ldots, x^*_T) = 0\) for \(t = 1, \ldots, T + 1\). (Remember that \(x^*_0 = x_0\) is given.) Hence, \((x^*_t, \ldots, x^*_T)\) must satisfy the Euler equation
\[
F_t^*(t, x_t, x_{t+1}) + F_t^*(t - 1, x_{t-1}, x_t) = 0, \quad t = 1, \ldots, T
\]
(3)
\[
F_t^*(t - 1, x_{t-1}, x_t) = 0, \quad t = T + 1
\]
(If \(x_{T+1}\) does not appear explicitly in \(F(T, x_T, x_{T+1})\), the last equation becomes trivial.)

This is a second-order difference equation analogous to the Euler equation in the classical calculus of variations. (See Section 8.2.) Note carefully that the partial derivatives in (3) are evaluated at different triples.
EXAMPLE 1: Consider the problem

\[
\max \left\{ \sum_{t=0}^{T-1} \ln c_t + \ln x_T \right\} \quad \text{subject to} \quad x_{t+1} = \beta (x_t - c_t) \quad \text{for} \quad t = 0, \ldots, T-1
\]

Here \( x_t \) is wealth at time \( t \), with \( x_0 \) fixed. An amount \( c_t \) is 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 control variables is

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

For \( t = T \), the Euler equation is \( F'_T(x_T, x_{T-1}, x_{T-2}) + F_T(t-1, x_{T-1}, x_T) = 0 \), with \( F_T(t, x_T, x_{T-1}) = \ln x_T \) and \( F(t-1, x_{T-1}, x_T) = \ln(x_{T-1} - \beta x_T) \). Therefore, the Euler equation reduces to \( 1/x_T - \beta (x_{T-1} - \beta x_T) = 0 \), or \( x_{T-1} - 2 \beta x_T \).

For \( t = 1, 2, \ldots, T-1 \), the Euler equation 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 \( x_{T-1} = 2\beta x_{T-2} - \beta^2 x_T = 4\beta x_T - \beta^2 x_T = 3\beta x_T \). More generally, given \( x_T \) and \( x_{T-1} = 2\beta x_T \), we can show by backwards induction that \( x_t = (T+1-t)\beta^{T-1} x_T \). This implies that \( x_0 = (T+1)\beta^T x_T \), so \( x_T = x_0 \beta^{-T} (T+1) \).

We conclude that the optimal solution of the problem is

\[
x_T^* = \frac{T + 1 - t}{T + 1} x_0 \beta^{-T}, \quad c_T^* = x_T^* - \beta x_{T+1} = \frac{\beta x_0 - x_T^*}{T + 1}
\]

We see that optimal consumption is steadily decreasing as \( T \) increases.

**NOTE 1** In Example 1 some might prefer to equate the partial derivatives of the maximand in (a) to 0 directly, rather than introducing the function \( F \). In particular, equating the partial derivative w.r.t. \( x_T \) to 0 yields \( -\beta (x_{T-1} - \beta x_T) + 1/x_T = 0 \); equating each partial derivative w.r.t. \( x_t \) to 0 yields \( -\beta (x_{t-1} - \beta x_t) + 1/(x_{t-1} - \beta x_{t+1}) = 0 \) for \( t = 1, 2, \ldots, T-1 \); etc.

**NOTE 2** Here is a general solution procedure for problem (1), similar to that used in Section 12.1. First, for \( t = T \) and for each fixed \( x_T \), find \( x^*_{T+1}(x_T) \) to maximize \( F(T, x_T, x_{T-1}) \); the associated first-order condition is \( F'_T(x_T, x_{T-1}, x_{T-2}) = 0 \), the appropriate version of (3). Next, \( x_{T+1} = x^*_{T+1}(x_T) \) is inserted into \( F(T, x_T, x_{T-1}) + F(t+1, x_{T+1}, x_T) \) for \( t = T - 1 \), and this expression is maximized w.r.t. \( x_T \), yielding \( x^*_{T+1}(x_{T-1}) \), using the first-order condition (3) for \( t = T - 1 \). Then \( x_T = x^*_{T+1}(x_{T-1}) \) is inserted into the expression

\[
F(t, x_t, x_{t+1}) + F(t+1, x_{t+1}, x_{t+2}) \quad \text{for} \quad t = T - 2, \quad \text{and this expression is maximized w.r.t.} \quad x_{t+1}, \quad \text{yielding} \quad x^*_{t+1}(x_{t-1}, x_t), \quad \text{using} \quad \text{(3) at} \quad t = T - 2.
\]

We continue to work backwards in this manner for \( t = T, T - 1, \ldots, 2, 1 \), until the last step we construct the function \( x^*_1(x_0) \). Since \( x_0 \) is given, we can work forward again to determine first \( x_1 = x^*_1(x_0) \), then \( x_2 = x^*_2(x_1) \), and so on. (In the example above we used another approach.)

**PROBLEMS FOR SECTION 12.2**

1. (a) Transform Problem 12.1.1 to the form (1).
   (b) Derive the corresponding Euler equation, and find its solution. Compare with the answer to Problem 12.1.1.

2. (a) Transform the problem in Example 12.1.3 to the form (1).
   (b) Derive the corresponding Euler equation, and find its solution. Compare with the answer in Example 12.1.3.

**12.3 Infinite Horizon**

Economists often study dynamic optimization problems over an infinite horizon. This avoids specifying 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 (12.2):

\[
\max \sum_{t=0}^{\infty} r^t f(x_t, u_t), \quad x_{t+1} = g(x_t, u_t), \quad t = 0, 1, 2, \ldots, x_0 \text{ given}, \quad u_t \in U \subseteq R \quad (1)
\]

Here \( f \) and \( g \) are given functions of two variables, there is a constant discount factor \( \beta \in (0, 1) \), and \( x_0 \) is a given number in \( R \). Having \( \beta \in (0, 1) \) is essential for the subsequent analysis of the problem in this section. Note that, apart from replacing the horizon \( T \) by \( \infty \) as the upper limit of the sum, the two functions \( f(t, x_t, u_t) \) and \( g(t, x_t, u_t) \) in (12.2) have been replaced by \( \beta f(x_t, u_t) \) and \( g(x_t, u_t) \) respectively. Because neither the new function \( f \) nor \( g \) depends explicitly on \( t \), problem (1) is called autonomous or stationary.

The sequence pair \((x_t, u_t)\) is called admissible provided that each control satisfies \( u_t \in U \), the initial state \( x_0 \) has the given value, and the difference equation in (1) is satisfied for all \( t = 0, 1, 2, \ldots \).

For simplicity, we begin by assuming that \( f \) satisfies the boundedness condition

\[
M_1 \leq f(x, u) \leq M_2 \quad \text{for all} \quad (x, u) \in U \quad (2)
\]

where \( M_1 \) and \( M_2 \) are given numbers. Because \( 0 < \beta < 1 \), the sum in (1) 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 \( \pi_t = (u_t, u_{t+1}, \ldots) \), where \( u_t \in U \) for \( t = s, s + 1, \ldots \). The successive states generated by this control sequence are found by solving \( x_{t+1} = g(x_t, u_t) \), 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 \( \pi_t \), starting from state \( x \) at time \( s \) is

\[
V_x(x, \pi_t) = \sum_{t=s}^{\infty} \beta^t f(x_t, u_t) = \beta^s V^t(x, \pi_t)
\]

where

\[
V^t(x, \pi_t) = \sum_{t=s}^{\infty} \beta^{t-s} f(x_t, u_t)
\]  (3)

The difference between \( V_t \) and \( V^t \) is that in \( V_t \) all benefits from time \( s \) on are discounted back to the fixed initial time 0, whereas in \( V^t \) they are discounted back to the variable starting time \( s \). Now let

\[
J_t(x) = \max_{\pi_t} V_t(x, \pi_t) = \beta^t J^t(x), \quad \text{where} \quad J^t(x) = \max_{\pi_t} V^t(x, \pi_t)
\]  (4)

and where the maxima are taken over all sequences \( \pi_t = (u_t, u_{t+1}, \ldots) \) with \( u_{t+k} \in U \).

Thus, \( J_t(x) \) is the maximum total discounted utility (or benefit) that can be obtained over all the periods from \( t = s \) to \( \infty \), given that the system starts in state \( x \) at time \( t = s \). We call \( J_t(x) \) the (optimal) value function for problem (1).

We next claim that the function \( J^t(x) \) satisfies the important property

\[
J^0(x) = J^t(x)
\]  (5)

Intuitively, this equality is obvious. 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 \( t \). So, finding either \( J^t(x) = \max_{\pi_t} V^t(x, \pi_t) \) or \( J^0(x) = \max_{\pi_0} V^0(x, \pi_0) \) requires solving essentially the same optimization problem, which therefore gives the same maximum value in each case. A more precise argument for (5) is given in Note 3 below.

Equations (4) and (5) together imply that

\[
J_t(x) = \beta^t J^0(x), \quad s = 0, 1, \ldots
\]  (6)

Define

\[
J(x) = J_0(x) = J^0(x)
\]  (7)

From (6) it follows that if we know \( J_0(x) = J(x) \), then we know \( J_t(x) \) for all \( t \). The main result in this section is the following:

**Theorem 12.3.1. (Fundamental Equation for the Infinite Horizon)**

The value function \( J_0(x) = J(x) \) in (4) for problem (1) satisfies the equation

\[
J(x) = \max_{u \in U} \left[ f(x, u) + \beta J(g(x, u)) \right] \quad \text{(the Bellman equation)}
\]  (8)

---

1 The existence of this maximum is discussed later in Note 4.

---

A rough argument for (8) resembles the argument for Theorem 12.1.1: Suppose we start in state \( x \) at time \( t = 0 \). If we choose the control \( u \), the immediate reward is \( \beta^t f(x, u) = f(x, u) \), and at time \( t = 1 \) we move to state \( x_1 = g(x, u) \). Choosing an optimal control sequence from \( t = 1 \) 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 \( t = 0 \) is one that maximizes the sum \( f(x, u) + \beta J(g(x, u)) \). The maximum of this sum is therefore \( J(x) \).

We call (8) a "functional equation" because the unknown is the function \( J(x) \) that appears on both sides. Under the boundedness condition (2), together with the assumptions that the maximum in (8) is attained and that \( 0 < \beta < 1 \), equation (8) always has one and only one bounded solution \( J(x) \), 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 (8) is the optimal control, which is therefore independent of \( t \).

In general it is difficult to use equation (8) to find \( J(x) \). The reason is that maximizing the right-hand side of (8) requires the function \( J(x) \) to be known beforehand.

**Example 1.** Consider the infinite horizon analogue of problem (i) in Example 12.1.4 in the case when \( a_t = a \) for all \( t \), independent of \( t \). We also introduce a new control \( u \) defined by \( u = x \).

Thus, \( u \) is the proportion of wealth \( x \) that is spent in the current period.

The former constraints \( u \in (0, x) \) is then replaced by \( u \in (0, 1) \). So the problem becomes

\[
\max_{\pi_t} \sum_{t=0}^{\infty} \beta^t (u_t x_t)^{1-\gamma}, \quad x_{t+1} = a_t(1 - u_t) x_t, \quad t = 0, 1, \ldots \quad u_t \in (0, 1) \quad (i)
\]

where \( a_t \) and \( x_t \) are positive constants, \( \beta \in (0, 1), \gamma \in (0, 1) \), and \( \beta a_t^{1-\gamma} < 1 \). 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.

In the notation of problem (1), we have \( f(x, u) = (u x)^{1-\gamma} \) and \( g(x, u) = a(1 - u)x \). Equation (8) therefore yields

\[
J(x) = \max_{u \in (0, 1)} \left[ (ux)^{1-\gamma} + \beta J(a(1 - u)x) \right]
\]  (ii)

In the closely related problem in Example 12.1.4, the value function was proportional to \( x^{1-\gamma} \). A reasonable guess in the present case is that \( J(x) = k x^{1-\gamma} \) for some positive constant \( k \). We try this as a solution. Then, after cancelling the factor \( x^{1-\gamma} \), (ii) reduces to

\[
k = \max_{u \in (0, 1)} \left[ u^{1-\gamma} + \beta ku a^{1-\gamma}(1-u)^{1-\gamma} \right]
\]  (iii)

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

\[
\phi'(u) = (1 - \gamma) u^{-\gamma} - (1 - \gamma) \rho u^{\gamma - 1} = 0
\]

implying that \( u^{1-\gamma} = \rho u^{\gamma - 1} \). Raising each side to the power \(-1/\gamma\) and then solving for \( u \), we see that the maximum of \( \phi \) is attained at
Then equation (iii) implies that $k$ satisfies the equation

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

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

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

Note that $\rho$ increases with the discount factor $\beta$ and with the return $\delta$ to saving, as an economist would expect.

In this example the boundedness assumption (2) is not valid without a simple transformation. Note that $ax_0$ is the maximum wealth the consumer could have accumulated by time $t$ by saving nothing, i.e., if $u_t = 0$ for $s = 0, 1, \ldots, t - 1$. Now define the modified state variable $y_t = x_t/u_0^{1/\gamma}$, which is the proportion of this maximum wealth that remains. Obviously $y_0 = 1$, and $y_t$ satisfies the difference equation $y_{t+1} = (1 - \delta)y_t$, so $1 \geq y_1 \geq y_2 \cdots \geq y_{t+1} \geq \cdots \geq 0$. The new objective function is $\sum_{t=0}^{\infty} \beta^t (y_t u_t^{1/\gamma})$, where $\beta = \delta^{1/\gamma}$ and so $0 < \beta < 1$. The transformed Bellman equation is

$$J(y) = \max_{a \in [0, 1]} \left\{ (a x_0)^{1/\gamma} + \beta J(1 - y) \right\}$$

This is easily seen to have $J(y) = J(ax_0 y) = (1 - \delta)^{-\gamma} (ax_0 y)^{1 - \gamma}$ as a solution, with the same optimal control $v = 1 - \rho$.

The transformed problem satisfies the restricted boundedness condition in Note 2 below, because the modified state $y_t$ remains within the interval $[0, 1]$ for all $t$, and so $0 \leq (x_0 y_0)^{1/\gamma} \leq x_0^{1/\gamma}$ for all $t$ and for all $y_t$ in $[0, 1]$. Therefore the control $v$ defined in (iv) really is optimal and the transformed problem is solved. So is the original problem, of course.

**NOTE 1** As pointed out in Note 12.1.5, the same theory applies without change when $x, u, v,$ and $g$ are vector functions. Moreover, $U$ may depend on the state, $U = U(x)$ (but not explicitly on time).

**NOTE 2** It suffices to assume that condition (2) holds for all $x$ in $X(x_0) = \bigcup_{x_0} X(x_0)$, where $X(x_0)$ is defined in Note 12.1.3. The function $J(x)$ need only be defined on $X(x_0)$.

**NOTE 3** To show (5) more formally, let $x$ be any fixed state. First, consider any policy sequence $\pi_x = (x_0, x_1, \ldots)$ that starts at time $s$. Now define the corresponding sequence $\pi_x^0 = (x_0^0, x_1^0, \ldots)$ shifted earlier so that it starts at time $0$ instead of at time $s$. Thus, $x_0^0 = x_0, x_1^0 = x_{s+1},$ and generally $x_t^0 = x_{t+s}$, for $t = 0, 1, \ldots$ Then, given the same starting state $x$, if $x$ and $x_t^0$ denote the states reached at time $t$ by following $\pi_x$ and $\pi_x^0$ starting at times $s$ and $0$ respectively, a moment’s reflection leads to the conclusion that $x_0^0 = x_0 + f(x_0, u_0)$, and so $f(x_0, u_0) = f(x_{s+1}, u_{s+1})$ for $t = 0, 1, \ldots$. It follows from (3) that $V^0(x, \pi_x^0) = V^0(x, \pi_x)$. But every shifted admissible policy $\pi_x^0$ is also admissible at time $0$, so we can use (4) to obtain

$$J^0(x) = \max_{\pi_x^0} V^0(x, \pi_x^0) \geq \max_{\pi_x^0} V^0(x, \pi_x) = \max_{\pi_x} V^0(x, \pi_x) = J^*(x)$$

On the other hand, consider any policy sequence $\pi_x = (x_0, u_0, x_1, \ldots)$ that starts at time $0$, and let $\pi_x^0 = (x_0^0, u_0^0, x_1^0, \ldots)$ be the corresponding sequence shifted to a later time $s$ so that $x_s^0 = x_0, u_s^0 = u_0, x_{t+1}^0 = u_t, \text{ for } t = s, s+1, \ldots$ Again, given the same starting point $x$, the states $x_t$ and $x_t^0$ reached at time $t$ by following $\pi_x$ and $\pi_x^0$ starting at times $0$ and $s$ respectively will satisfy $x_t = x_t^0$, and so $f(x_t, u_t) = f(x_t^0, u_t^0)$ for $t = s, s+1, \ldots$. Then (3) and (4) imply that $V^0(x, \pi_x^0) = V^0(x, \pi_x^0)$, so

$$J^0(x) = \max_{\pi_x^0} V^0(x, \pi_x^0) = \max_{\pi_x^0} V^0(x, \pi_x^0) \leq \max_{\pi_x} V^0(x, \pi_x) = J^*(x)$$

From (9) and (10) we conclude that $J^0(x) = J^*(x)$, which is (5).

**NOTE 4** Whenever we wrote “max” above, it was implicitly assumed that the maximum exists. Of course, without further conditions on the system, this may not be true. Under the boundedness condition (2), the same assumptions as in the finite horizon case ($f$ and $g$ are continuous and $U$ is compact) ensure that the maxima in (4) and (8) do exist.

Many economic applications, however, do not satisfy the boundedness condition (2). So let us investigate what happens when we replace max with sup in (4), as well as when the set $U(x)$ depends on $x$, as in Note 1. In fact, suppose the sum $\sum_{t=0}^{\infty} \beta^t f(x_t, u_t)$ always exists (possibly with an infinite value), so $J_0(x_0) = \sup_{\pi_x} V_0(x_0, \pi_x)$ must exist. By the result (A.4.7) on iterated suprema, we have

$$J_0(x_0) = \sup_{u_0 \in U(x)} \left\{ f(x_0, u_0) + \sum_{t=1}^{\infty} \beta^t f(x_t, u_t) \right\} = \sup_{u_0 \in U(x)} \left\{ f(x_0, u_0) + \sum_{t=1}^{\infty} \beta^t f(x_t, u_t) \right\}$$

So the modification

$$J(x) = \sup_{u \in U(x)} \left\{ f(x, u) + \beta J(g(x, u)) \right\}$$

of the Bellman equation (8) still holds even if no maximum exists.

Next, let us use the contraction mapping theorem 14.3.1 to prove that version (11) of the Bellman equation has a unique solution.

Indeed, define the operator $T_0$ on the domain $\mathcal{B}$ of all bounded functions $I(x)$ so that

$$T_0(I)(x) = \sup_{u \in U(x)} \left\{ f(x, u) + \beta I(g(x, u)) \right\}$$
for all \( t \) and all \( x \). As in Section 14.3, the distance between any two bounded functions \( \bar{J} \) and \( J \) is defined as \( d(\bar{J}, J) = \sup_{x} |J(z) - \bar{J}(z)| \). Then

\[
T(J)(x) = \sup \left\{ \int f(x, u) + \beta \bar{J}(g(x, u)) + \beta(\bar{J}(g(x, u)) - J(g(x, u))) \right\} \\
\leq \sup \left\{ \int f(x, u) + \beta \bar{J}(g(x, u)) + \beta d(\bar{J}, J) \right\} = T(\bar{J})(x) + \beta d(\bar{J}, J)
\]

Symmetrically, \( T(J)(x) \leq T(\bar{J})(x) + \beta d(\bar{J}, J) \). So \( |T(J)(x) - T(\bar{J})(x)| \leq \beta d(\bar{J}, J) \), implying that

\[
d(T(J), T(\bar{J})) = \sup_{x} |T(J)(x) - T(\bar{J})(x)| \leq \beta d(\bar{J}, J) \quad (\ast\ast)\]

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

Finally, we check that any control \( u = \bar{u} \) that yields a maximum in the Bellman equation (11) is optimal. To see this, let \( T^{\bar{u}} \) be the operator on \( \mathcal{J} \) which is defined by \((a,s)\) when \( U(x) \) takes the form \((\bar{u}(s))\) (leaving no choice except \( u = \bar{u}(s) \)). By definition of \( \bar{u} \), the unique solution \( J \) of the Bellman equation also satisfies \( T^{\bar{u}}(J) = J \). Also, because \( T^{\bar{u}} \) satisfies \((a,s)\) for \( U(x) = (\bar{u}(s)) \), we have \( T^{\bar{u}}(J) = J \). But \( T^{\bar{u}} \), like \( T \) itself, is a contraction mapping, so \( T^{\bar{u}}(J) = J \) has a unique solution. It follows that \( J = J(\bar{u}) \), and in particular, the supremum \( J(x) \) for any \( x \) is equal to \( J(\bar{u})(x) \), the value attained by following the control policy \( \bar{u} \).

### PROBLEMS FOR SECTION 12.3

**Problem 1.** Consider the problem

\[
\max_{\alpha \in \mathbb{R}} \sum_{n=0}^{\infty} \beta^n (-\alpha e^{-\alpha} - \frac{1}{2} e^{-\alpha}) \quad x_{t+1} = x_t - \alpha u_t \quad t = 0, 1, \ldots, \quad x_0 \text{ given}
\]

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

**Problem 2.** (a) Consider the following problem with \( \beta \in (0, 1) \):

\[
\max_{\alpha \in \mathbb{R}} \sum_{n=0}^{\infty} \beta^n (-\frac{1}{2} x^2_t - u_t^2) \quad x_{t+1} = x_t + u_t, \quad t = 0, 1, \ldots, \quad x_0 \text{ given}
\]

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

(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 \( |u_t| \leq |u_{t-1}| \) for \( t = 1, 2, \ldots \). Does Note 2 then apply?

### 12.4 The Maximum Principle

Dynamic programming is the most frequently used method for solving discrete time dynamic optimization problems. An alternative solution technique is based on the so-called maximum principle. The actual calculations needed are often simpler. However, when there are terminal restrictions on the state variables, the maximum principle is often preferable. The corresponding principle for optimization problems in continuous time is studied in more detail in Chapters 9 and 10, because for such problems it is the most important method.

**Example 1.** Consider first the discrete time dynamic optimization problem with one state, one control variable and a free end state:

\[
\max_{u_t \in \mathbb{R}} \sum_{t=0}^{T} f(t, x_t, u_t) \quad x_{t+1} = g(t, x_t, u_t), \quad t = 0, \ldots, T - 1, \quad x_0 \text{ given}, \quad x_T \text{ free (1)}
\]

Here we assume that the control region \( U \) is convex, i.e. an interval. The state variable \( x_t \) evolves from the initial state \( x_0 \) according to the law of motion in (1), with \( u_t \) as a control that is chosen at each \( t = 0, \ldots, T \). Define the Hamiltonian by

\[
H(t, x, u, p) = \begin{cases} f(t, x, u) + pg(t, x, u) & \text{for } t < T \\ f(t, x, u) & \text{for } t = T \end{cases}
\]

where \( p \) is called an adjoint variable (or co-state variable).

**Theorem 12.4.1 (The Maximum Principle: Necessary Conditions):**

Suppose \((x^*_t, u^*_t)\) is an optimal sequence pair for problem (1), and let \( H \) be defined by (2). Then there exist numbers \( p_t \), with \( p_T = 0 \), such that for all \( t = 0, \ldots, T \)

\[
H_t^n(t, x^*_t, u^*_t, p_t) = 0 \quad \text{for all } u \in U
\]

(3)

(Note that if \( u^*_t \) is an interior point of \( U \), (3) implies that \( H^n_t(t, x^*_t, u^*_t, p_t) = 0 \).) Furthermore, \( p_t \) is a solution of the difference equation

\[
p_{t-1} = H^n_t(t, x^*_t, u^*_t, p_t) \quad t = 1, \ldots, T
\]

(4)

**Note 1.** In Theorem 12.4.1 there are no terminal conditions. When there are terminal conditions, Theorem 12.5.1 gives necessary conditions for the case of several variables. For a proof of these two theorems see Arkin and Evtinheen (1987). A closer analogy with the continuous time maximum principle comes from writing the equation of motion as \( x_{t+1} = g(t, x_t, u_t) \). If we redefine the Hamiltonian accordingly, then (4) is replaced by \( p_t - p_{t-1} = -H^n_t(t, x^*_t, u^*_t, p_t) \), which corresponds to equation (9.2.5).

Sufficient conditions are given in the following theorem. The proof is similar to the proof of the corresponding theorem in continuous time.
THEOREM 12.4.2 (SUFFICIENT CONDITIONS).

Suppose that the sequence triple \((x^*_r, u^*_r, \{p_t\})\) satisfies all the conditions in Theorem 12.4.1, and suppose further that \(H(t, x, u, p_t)\) is concave with respect to \((x, u)\) for every \(r\). Then the sequence triple \((x^*_r, u^*_r, \{p_t\})\) is optimal.

NOTE 2 Suppose that admissible pairs are also required to satisfy the constraints \((x_t, u_t) \in A_t, t = 0, 1, \ldots, T\), where \(A_t\) is a convex set for all \(t\). Then Theorem 12.4.2 is still valid, and \(H\) need only be concave in \(A_t\).

NOTE 3 If \(U\) is compact and the functions \(f\) and \(g\) are continuous, there will always exist an optimal solution. (This result can be proved by using the extreme value theorem.)

NOTE 4 A weaker sufficient condition for optimality than in Theorem 12.4.2 is that for each \(r\) the pair \((x^*_r, u^*_r)\) maximizes \(H(t, x, u, p_t) - p_{t+1}x\) as a function of \(u \in U\) and \(x \in R\).

EXAMPLE 2 Apply Theorem 12.4.2 to the problem in Example 12.1.2,

\[
\max \sum_{t=0}^{3} (1 + x_t - u^2_t) \quad x_{t+1} = x_t + u_t, \quad x_0 = 0, \quad u_t \in R, \quad t = 0, 1, 2, 3.
\]

Solution: For \(r < 3\), the Hamiltonian is \(H = 1 + x - 2u + p(x + u)\), so \(H^*_r = x - 2u + p = -2\). For \(r = 3\), \(H = 1 - u^2\), so \(H^*_r = -2u + p = 1\). Note that the Hamiltonian is concave in \((x, u)\). The control region is open, so \(H^*_0 = 0\), i.e., \(-2u_t^2 + p_t = 0\) for \(t = 0, 1, 2\), and \(-2u^2_t = 0\) for \(t = 3\). Thus \(u^*_0 = \frac{1}{2}p_0, u^*_1 = \frac{1}{2}p_1, u^*_2 = \frac{1}{2}p_2, \) and \(u^*_3 = \frac{1}{2}p_3\).

The difference equation (4) for \(p_t\) is \(p_{t+1} = 2 + p_t, t = 1, 2\), and so \(p_0 = 1 + p_1, p_1 = 1 + p_2, p_2 = 1 + p_3\). For \(t = 3\), (2) yields \(p_3 = 1\), and we know from Theorem 12.4.1 that \(p_3 = 0\).

It follows that \(p_2 = 1, p_1 = 1 + p_2 = 2, p_0 = 1 + p_1 = 3\). This implies the optimal choices \(u^*_0 = 1, u^*_1 = 1, u^*_2 = 1/2\), and \(u^*_3 = 0\) for the controls, which is the same result as in Example 12.1.2.

EXAMPLE 3 Consider an oil field in which \(x_0 > 0\) units of extractable oil remain at time \(t = 0\). Let \(u_t \geq 0\) be the rate of extraction and let \(x_t\) be the remaining stock at time \(t\). Then \(u_t = x_t - x_{t+1}\). Let \(C(t, x_t, u_t)\) denote the cost of extracting \(u_t\) units in period \(t\) when the stock is \(x_t\). Let \(w\) be the price per unit of oil and let \(r\) be the discount rate, with \(\beta = 1/(1 + r) \in (0, 1)\) the corresponding discount factor. If \(T\) is the fixed end of the planning period, the problem of maximizing total discounted profit can be written as

\[
\max_{u_t \geq 0} \sum_{t=0}^{T} [w u_t - C(t, x_t, u_t)], \quad x_{t+1} = x_t - u_t, \quad t = 0, 1, \ldots, T - 1, \quad x_0 > 0
\]

assuming also that
\(u_t \leq x_t, \quad t = 0, 1, \ldots, T\),

because the amount extracted cannot exceed the stock.

Because of restriction (ii), this is not a dynamic optimization problem of the type described by (1). However, if we define a new control \(v_t = u_t + v_t\), then restriction (iii) combined with \(v_t \geq 0\) reduces to the control restriction \(v_t \in [0, 1]\), and we have a standard dynamic optimization problem. Assuming that \(C(t, x, v) = w^2 - x\), and \(0 < w < 1\), apply the maximum principle to find the only possible solution of the problem

\[
\max_{u_t \geq 0} \sum_{t=0}^{T} [w u_t - C(t, x_t, u_t)], \quad x_{t+1} = x_t(1 - v_t), \quad x_0 > 0, \quad v_t \in [0, 1]
\]

with \(x_T\) free.

Solution: We denote the adjoint function by \(p_t\). We know that \(p_T = 0\). The Hamiltonian is \(H = \beta^T(w u - w^2 x) + px(1 - v)\). (This is valid also for \(t = T\), because then \(p_T = 0\)).

Then \(H_T = \beta^T(w u - 2w x) - px\) and \(H_T' = \beta^T(w u - v) + p(1 - v)\). So (3) implies that, for \((v^*_t, \{p_t\})\) to solve the problem, there must exist numbers \(p_t\), with \(p_T = 0\), such that, for all \(t = 0, 1, \ldots, T\),

\[
[b^T u^*_t (w - 2u^*_t) - p_t x^*_t] (v - v^*_t) \leq 0 \quad \text{for all} \quad v \in [0, 1]
\]

For \(t = T\), with \(p_T = 0\), this condition reduces to

\[
[b^T u^*_T (w - 2u^*_T) - p_T x^*_T] (v - v^*_T) \leq 0 \quad \text{for all} \quad v \in [0, 1]
\]

Having \(v^*_T = 0\) would imply that \(u_T \leq 0\) for all \(v \in [0, 1]\), which is impossible because \(w > 0\). Suppose instead that \(v^*_T = 1\). Then (v) reduces to \(b^T u^*_T (w - 2u^*_T) - p_T x^*_T \geq 0\) for all \(v \in [0, 1]\), which is impossible because \(w < 2 < 0\) (put \(v = 0\)). Hence, \(v^*_T \in (0, 1)\). For \(t = T\), condition (v) then reduces to \(b^T u^*_T (w - 2u^*_T) = 0\), and so

\[
v^*_T = \frac{1}{2} w
\]

According to (4), for \(t = 1, \ldots, T\),

\[
p_{t-1} = b^T u^*_t (w - v^*_t) + p_1(1 - v^*_t)
\]

For \(t = T\), because \(p_T = 0\) and \(v^*_T = \frac{1}{2} w\), this equation reduces to

\[
p_{T-1} = b^T u^*_T (w - v^*_T) = \frac{1}{2} w^2 \beta T
\]

For \(t = T - 1\), the expression within square brackets in (iv) is

\[
b^T u^*_T (w - 2u^*_T) - p_{T-1} x^*_T (1 - v^*_T) = b^T x^*_T (w - 1 - \beta w - 2v^*_T)\]

Because \(0 < w < 1\) and \(\beta \in (0, 1)\), one has \(1 > \frac{1}{2} w\). It follows that both \(v^*_T = 0\) and \(v^*_T = 1\) are impossible as optimal choices in (iv). So the maximizer \(v^*_T \in (0, 1)\) in (iv) must be interior, implying that the square brackets in the last line of (iv) is 0. Hence,

\[
v^*_T = \frac{1}{2} w (1 - \frac{1}{2} \beta w)
\]
Let us now go $k$ periods backwards in time. Define $q_{T-k} = \beta q_{T-k+1}$. We prove by backward induction that at each time $T - k$ we have an interior maximum point $u^*_{T-k}$ in (iv). Then $u^*_{T-k} = \frac{q_{T-k}}{\beta}$, which belongs to $(0,1)$ if $q_{T-k} \in (0,2)$. We prove by backward induction that $q_{T-k} \in [0,w]$. Suppose this is valid for $k$. Let us show it for $k + 1$. Using (vii) and the definition of $q_{T-k}$, we find that $q_{T-(k+1)} = F(q_{T-k})$ where $F(q) = \frac{1}{4}(w - q)^2 + q$. Note that $q \mapsto F(q)$ is a strictly convex function and, by the assumptions on the parameters, we have $0 < F(q) \leq \max(F(0), F(w)) = \max(\frac{w^2}{4}, w) < w$ for all $q \in [0,w]$. Hence, $q_{T-(k+1)} \in [0,w]$. Because $q_T = 0$, the backward induction can be started. Thus the solution of the problem is given by $u^*_{T-k} = (1/2)(w - q_{T-k})$, where $q_{T-k}$ is determined by $q_{T-(k+1)} = \beta^{-1}q_{T-k} = F(q_{T-k})$ with $q_T = 0$.

PROBLEMS FOR SECTION 12.4

1. Consider Problem 12.1.1. Write down the Hamiltonian, condition (1), and the difference equation for $p_t$. Use the maximum principle to find a unique solution candidate. Verify that the conditions in Theorem 12.4.2 are satisfied, and that you have found the optimum solution.

2. (Boltyanskii) Consider the problem

$$\max_{u_t \in (0,1)} \sum_{i=a}^T (u_i^2 - 2x_i^2) \text{ s.t. } x_{i+1} = u_i, \quad t = 0, 1, \ldots, T - 1, \quad x_0 = 0$$

(a) Prove that $u_i^* = 0$ for $t = 0, 1, \ldots, T - 1$, and $u_T^* = 1$ or $-1$ are optimal controls. (Express the objective function as a function of $u_0, u_1, \ldots, u_T$ only.)

(b) Verify that, although the conditions in Theorem 12.4.2 are satisfied, $u_T^*$ does not maximize $H(t, x_t, u, p_t)$ subject to $u \in [-1, 1]$.

12.5 More Variables

Consider the following end constrained problem with $n$ state and $r$ control variables:

$$\max \sum_{i=0}^n f_i(x_i, u_i), \quad x_{i+1} = g_i(x_i, u_i), \quad x_0 \text{ is given}, \quad u_i \in U \subseteq \mathbb{R}^r$$

Here $x_i$ is a state vector in $\mathbb{R}^n$ that evolves from the initial state $x_0$ according to the law of motion in (1), with $u_i$ as a control vector in $U$ that is chosen at each $t = 0, \ldots, T$. We put $x_0 = (x_0^1, \ldots, x_0^n), u_0 = (u_0^1, \ldots, u_0^n)$, and $g = (g_1^1, \ldots, g_1^n)$. We assume that the control region $U$ is convex.

The terminal conditions are assumed to be

(a) $x_T^i = \tilde{x}^i$ for $i = 1, \ldots, l$

(b) $x_T^i \geq \tilde{x}^i$ for $i = l + 1, \ldots, m$

(c) $x_T^i$ free for $t = m + 1, \ldots, n$

where $0 \leq l \leq m \leq n$. Define the Hamiltonian by

$$H(t, x, u, p) = \left\{ \begin{array}{ll}
q_0 f_i(t, x, u) + \sum_{i=1}^r p_i^t g_i^t(t, x, u) & \text{for } t < T \\
f_i(t, x, u) & \text{for } t = T
\end{array} \right.$$
The economic interpretation is that a consumer wants to maximize a sum of discounted utilities $β^t U(u_t)$ up to a fixed horizon $T$. We assume that $U'(u_t) > 0$ and $U''(u_t) < 0$ for $u_t > 0$. The coefficient $β$ is the subjective discount rate. Wealth $x_t$ develops according to the given difference equation, where $γ_t$ is income at time $t$, and $α_t = (1 + r)$ with $r$ as the interest rate. (A consumer who deposits $x_t + γ_t - u_t$ in an interest-bearing account at time $t$ receives $x_{t+1}$ at time $t + 1$.)

**Solution:** The Hamiltonian is $H = H(t, x_t, u_t, p_t) = β^t U(u_t) + p_t (x_t + γ_t - u_t)$ for $t = 0, 1, ..., T - 1$, and $H = H(T, x_T, u_T, p_T) = 0$ for $t = T$. Clearly $H$ is concave in $(x_t, u_t)$, so we use sufficient conditions with $q_0 = 1$. According to (4) and (5) we get $p_{t+1} = α_p p_t$ for $t < T$ and $p_T = p_T$. It follows that for $t < T$ we obtain $p_t = α^t p_0$, so $p_t ≥ 0$. Because the control region is open, condition (3) reduces to $H'_t(t, x_t^*, u_t^*, p_t) = 0$ for $t < T$ (see Note 2). This means that $β^t U'(u_t^*) = α^t p_t α^t - 1 = 0$, so $U'(u_t^*) = p_t α^t (α β)^{-1}$. In particular, optimality requires

$$\frac{U'(u_t^*)}{U''(u_{t+1}^*)} = α β$$

Thus, the ratio of the marginal utilities from one period to the next is the constant $α β$, equal to the discounted one-period rate of return. Note that consumption is constant in case $α β = 1$, rising in case $α β > 1$, and falling in case $α β < 1$.

Because $U'$ is strictly decreasing and so has an inverse, $u_t^* = (U')^{-1}(p_t α^t (α β)^{-1})$. In particular we see that $p_0 ≠ 0$, so $p_t > 0$. Then (6)(b) implies that $x_t^* = 0$. But $x_t^* = α^t x_0 + \sum_{i=1}^{t-1} α^{t-1-i}(γ_{t-1} - u_{t-1}^*) = α^t x_0 + \sum_{i=1}^{T-1} α^{T-1-i}(γ_{t-1} - (U')^{-1}(p_t α^t (α β)^{-1}))$, using formula (11.1.8). The equality $x_T^* = 0$ can then be used to determine $p_T$ uniquely.

### Infinite Horizon

We consider briefly the following infinite horizon version of problem (1)–(2):

$$\max_{u_t ∈ U} \sum_{t=0}^{∞} f(t, x_t, u_t), \quad x_t ∈ \mathbb{R}^n, \quad u_t ∈ U ⊆ \mathbb{R}^k, \quad U \text{ convex}$$

(7)

where we maximize over all sequence pairs $(x_t, u_t)$, satisfying

$$x_{t+1} = g(t, x_t, u_t), \quad t = 1, 2, ..., \quad x_0 \text{ given}$$

(8)

and the terminal conditions:

(a) $\lim_{t → ∞} x_t(T) = \tilde{x}_i, \quad i = 1, ..., l$

(b) $\lim_{t → ∞} x_t(T) ≥ \tilde{x}_i, \quad i = l + 1, ..., m$

(c) no condition, $\quad i = m + 1, ..., n$

Note that $f$ and $g = (g_1, ..., g_n)$ can now depend explicitly on $t$. Assume that the sum in (7) exists for all admissible sequences. The functions $f$ and $g$ are assumed to be $C^1$ with respect to all $x_t$ and $u_t$.

---

### PROBLEMS FOR SECTION 12.5

1. Consider the problem

$$\max_{u_t ∈ U} \sum_{t=0}^{∞} [1 + x_t - γ_t - 2u_t - u_t^2] \quad \text{s.t.} \quad \begin{cases} x_{t+1} = x_t - u_t, \quad x_0 = 5 \\ y_{t+1} = x_t + u_t, \quad y_0 = 2 \end{cases}, \quad t = 0, 1$$

(a) Solve the problem by using the difference equations to express the objective function $J$ as a function of only $x_0, u_1, u_2, \ldots, u_t$, and then optimize.

(b) Solve the problem by using dynamic programming. (Hint: Find $J_2(x, y), J_1(x, y)$, and $J_0(x, y)$ and the corresponding optimal controls.)

(c) Solve the problem by using Theorem 12.5.1.

2. Solve the problem

$$\max_{u_t ∈ U} \sum_{t=0}^{∞} (-x_t^2 - u_t^2) \quad \text{subject to} \quad x_{t+1} = x_t, \quad y_{t+1} = y_t + u_t, \quad t = 0, 1, ..., T = 1$$

where $x_0 = x_0$ and $y_0 = y_0$ are given numbers and $u_t ∈ U$.

3. Solve the problem

$$\max_{u_t ∈ U} \sum_{t=0}^{∞} β^t \ln(x_t - u_t) \quad \text{subject to} \quad x_{t+1} = u_t, \quad x_0 > 0, \quad u_t > 0$$

where $β ∈ (0, 1)$. Verify that $x_t^* > u_t^*$ for all $t$. 

---

4 See Section 10.3 or A.3 for the definition of $\lim$ or $\liminf$. 

---

**THEOREM 12.5.2 (SUFFICIENT CONDITIONS).**

Suppose that the sequence triple $(x_t^*, u_t^*, p_t)$ satisfies the conditions (3)–(4) with $q_0 = 1$. Suppose further that $U$ is convex and the Hamiltonian $H(t, x_t, u_t, p_t)$ is concave in $(x_t, u_t)$ for every $t$. Then $(x_t^*, u_t^*)$ is optimal provided that the following transversality condition is satisfied: for all admissible sequence pairs $(x_t, u_t)$,

$$\lim_{t → ∞} p_t(x_t - x_t^*) ≥ 0$$

(10)

**NOTE 3** Suppose that any admissible sequence $(x_t, u_t)$ is required to satisfy additional constraints. Then (10) needs only to be tested for such sequences.
interest payment is $0.07 = 585.641(1 - (0.07)^{-1}) = 527.20$, and so the principal repayment is $8058.64 - 527.20 = 7331.44.

3. (a) Let the remaining debt on January in year $n$ be $D_n$. Argue why $D_{n-1} - D_n = \frac{3}{2}D_{n-1}$, $n = 1, \ldots$, and $(1 - r/2)^{1/2} = (1/2)D$ implies that $r = 2 - 2.21090 \approx 0.139394$. (c) See SM.

4. See SM.

11.3

1. $(a) x_{t+1} = A + Bx_t + \varepsilon_t + \mu_t$. The characteristic polynomial is $(1 - \rho_1) = 1 - B$. (b) $x_{t+1} = A + Bx_t + \varepsilon_t + \mu_t$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$.

2. $x_t = A + Bt + \varepsilon_t$ is a solution: $x_{t+1} = x_t + \varepsilon_{t+1}$, so $x_{t+1} = A + B(t + 1) + \varepsilon_{t+1}$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$.

3. $x_t = A + Bt + \varepsilon_t$ is a solution: $x_{t+1} = x_t + \varepsilon_{t+1}$, so $x_{t+1} = A + B(t + 1) + \varepsilon_{t+1}$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$.

4. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$. (b) $x_t = A + Bt + \varepsilon_t$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$.

5. Both characteristic roots complex: $\beta \alpha < 4(1 - \alpha^2)$. Stability: $1 + \alpha^2 \beta < 4u + \alpha < 1$.

11.6

1. $(a) x_t = A + Bt + \varepsilon_t$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$. (b) $x_t = A + Bt + \varepsilon_t$. The characteristic polynomial is $(1 - \rho_1)^2 = 1 - 2B + B^2$.

2. For $\lambda = \sqrt{B}x_t = \lambda((A - B) + \varepsilon_t)$, $y_t = \frac{1}{\lambda}((A - B) - y_t)$. (b) $x_t = (A + Bt + \varepsilon_t)$, $y_t = \frac{1}{\lambda}((A - B) + \varepsilon_t)$. (c) $y_t = (A - B) + \varepsilon_t$.

3. $(a) y_t = 0.5292, 0.1839, y_{t-1} = 0$. (b) Solutions of the characteristic equation: $m_1 \approx 0.61, m_2 \approx 0.31$. Thus $y_t \approx 0.61 y_{t-1} + 0.31 y_{t-2} + 0.17 y_{t-3}$.

11.7

1. Let $g(x) = f(x) - \lambda x$. Then $g(\xi_1) = \xi_1 - \xi_2$ and $g(\xi_2) = \xi_1 - \xi_2$ have opposite signs, and the intermediate value theorem implies that $g(x)$ has a zero at some point between $\xi_1$ and $\xi_2$. $x^* = x^*$ is an equilibrium.

2. $x = -2.95735$. (b) $x = -2x - x \times \ln(x + 3)$, the positive root of $f(x) = 0$ is $x = 0$.

3. The cycle points are $\xi_1 = (25 - 3\sqrt{5})/10 \approx 1.82918, \xi_2 = (25 + 3\sqrt{5})/10 \approx 3.17082$. Since $f'(\xi_1) > 0$, the cycle is locally asymptotically stable. The equilibrium states are $x_1 = (15 + 3\sqrt{5})/10 \approx 1.50524, x_2 = (15 - 3\sqrt{5})/10 \approx 0.29584$, both equilibria are locally unstable. It is also clear that $x_1$ lies between $\xi_1$ and $\xi_2$.

Chapter 12

12.1

1. $x_2(1) = x_1 - 2$. For $x_2(0) = 0, x_1(0) = 2$, it follows that $x_2(3) = 2 - x_2(1)^2/3$ for $x_1(t) = x_2(t)$.

2. $x_2(1) = x_1 - 2$. For $x_2(0) = 0, x_1(0) = 2$, it follows that $x_2(3) = 2 - x_2(1)^2/3$ for $x_1(t) = x_2(t)$.
3. (a) With $\beta = (1 + r)^{-1}$, $J(x) = \beta^T x$ for $u(x)^T = 1$. For $\beta < 1$, $J(x) = \beta^T x$ with $u(x)^T = 1$; for $\beta \geq 1$, $J(x) = \beta^T x$ with $u(x)^T = 0$. (b) For $\beta < 1$, $P_\beta = \beta^2 P_1$; for $\beta \geq 1$, $P_\beta = \beta^{-2} P_1$. For $\beta < 1$, $J(x) = \beta^{-2} x$ and $u(x) = \beta^{-2} x$. For $\beta \geq 1$, $J(x) = \beta^{-2} x$ and $u(x) = 0$. (c) $\beta = 1$, $\gamma = 1$.

4. (a) and (b) $J(x) = (2n^2 + 3c^2) x^T$ with $x^T > 0$ and $x^T = 1$ for $n = 1, \ldots, T$.

5. $J(x) = \ln x$, $x^T > 0$ is arbitrary, $J(x) = \ln x + C$ with $C = \ln(3/2) - 1/3$. The optimal controls are $u(x)^T = 1$, $v(x)^T = 1$, $w(x)^T = 1$. The solution of this problem is $x_T = \beta^{-2} x_0$.

6. $J(x) = \max_{a \in A} (x^T - ax^T) = x^T \min_{a \in A} \{a \}$. The solution is $x_T = \beta^{-2} x_0$.

7. $J(x) = \max_{a \in A} (x^T - ax^T) = x^T \min_{a \in A} \{a \}$. The solution is $x_T = \beta^{-2} x_0$.

8. $J(x) = \max_{a \in A} (x^T - ax^T) = x^T \min_{a \in A} \{a \}$. The solution is $x_T = \beta^{-2} x_0$.

12.2

1. (a) $F(t, x, \alpha_0) = 1 - x^2 - 2(1 - x^2)T$ for $t = 0$. $F(2, x, T) = 1 - x^2$.

2. The Euler equation gives $x^T = (x - 2)T \alpha_0 + 1$, which yields the $x^T = (x - 2)T \alpha_0$ for $t = 1$. With $x_0 = 5$, $x = 30/11$ and $x_T = 20/11$, as in Problem 12.1.1.

12.3

1. Inserting $J(x) = -\alpha x^T$ into the Bellman equation yields:

$$J(x) = \max_{u(x)^T = 1} \{ -\alpha x^T \} = -\alpha x^T - \alpha x^T e^{-2 x^T}.$$ The maximizing $u(x)^T$ is $u(x)^T = 1 - \ln(x^T)$, and the equation reduces to $a = 2\alpha^2 x^T / 3$. Then $J(x) = \beta x^T + 1/3$, and so $u(x)^T = 3 x^T + 3$. For optimality, see SM.

2. (a) $J(x) = \alpha x^T$ is $3 x^T$, and the only positive solution is $x^T = 3 x^T$. (b) Note 2 applies for $x^T = x^T$.

12.4

$H = 1 - (x^T + 2a^T) + p(x - u)$ for $t = 0$, 1, and $H = 1 - (x^T + 2a^T)$ for $t = 2$. Condition (3) yields $p_0 = -4a^T$ and $p_1 = -4a^T$, and condition (4) gives $p_0 = -2x^T + p_1 = -2x^T$. Now, $x^T = x^T - u^T = 2a^T + 2a^T = 3 x^T - 2a^T$.

12.5

1. (a) $I = 3x + 3a^T - 3x^T + 3a^T x^T = 2a^T x^T + 3a^T x^T - 3 x^T - 3 x^T = -2x^T + \alpha x^T - \alpha x^T$.

2. The Hamiltonian is $H = -a x^T + p U + q(x + u)$ for $t = 0$, $H = -a x^T$ for $t = 1$, and $H = -a x^T$ for $t = 2$. The conditions of Theorem 12.5.1, using Note 3, are therefore sufficient:

3. (a) $p_0 = 0$, $p_1 = 0$, $p_2 = 0$, $p_3 = 0$, $p_4 = 0$.

4. $J(x) = \alpha x^T + 2 \alpha x^T$.

5. $J(x) = \alpha x^T + 2 \alpha x^T$.

12.6

1. $J(x) = \alpha x^T + 2 \alpha x^T$.

2. $J(x) = \alpha x^T + 2 \alpha x^T$.

3. $J(x) = \alpha x^T + 2 \alpha x^T$.
Solution: The objective function is $F(t, A, \dot{A}) = U(rA + w - \dot{A})e^{-rt}$. The Euler equation is easily shown to be

$$
\dot{A} - rA + (\rho - r)U' = 0
$$

($)\hspace{1cm}$(*)

Because $U' > 0$, one has $\delta F / \delta \dot{A} = -U''(C)e^{-rt} < 0$ everywhere. Therefore (3) implies that $A^*(T) = A_T$. Hence, any optimal solution $A^*(t)$ of the problem must satisfy (e) with $A^*(0) = A_0$ and $A^*(T) = A_T$. Because of the requirement imposed on $U$, the function $F(t, A, \dot{A})$ is concave in $(A, \dot{A})$ (for the same reason as in Example 8.4.1). Note that we have not proved that (e) really has a solution that satisfies the boundary conditions. See Problem 4 for a special case.

**PROBLEMS FOR SECTION 8.5**

1. Solve the problem

$$
\min \int_0^T (\dot{x}^2 + x^2) dt, \quad x(0) = 1, \quad (i) \text{ with } x(1) \text{ free, } (ii) \text{ with } x(1) \geq 1
$$

2. (a) Solve the variational problem

$$
\max \int_0^T (10 - \dot{x}^2 - 2\dot{x}e^{-rt} - 5x^2) e^{-rt} dt, \quad x(0) = 0, \quad x(1) = 1
$$

(b) What is the optimal solution if (i) $x(1)$ free? (ii) $x(1) \geq 1$?

3. J.K. Sengupta has considered the problem

$$
\min \int_0^T (a_1\dot{y}^2 + a_2G^2) dt, \quad \ddot{y} = r_1\dot{y} - r_2G, \quad \ddot{y}(0) = Y_0, \quad \ddot{y}(T) \text{ free}
$$

where $a_1, a_2, r_1, r_2, T,$ and $Y_0$ are given positive constants. Formulate this as a variational problem with $\ddot{y} = \ddot{y}'$ as the unknown function. Find the corresponding Euler equation, and solve the problem.

4. Solve the problem in Example 3 when $U(C) = -\alpha e^{-\beta C}$, with $\alpha > 0$ and $\beta > 0$.

5. (a) A community wants to plant trees to cover a 1500 hectare piece of land over a period of 5 years. Let $x(t)$ be the number of hectares that have been planted by time $t$, and let $u(t)$ be the rate of planting, $\dot{x}(t) = u(t)$. Let the cost per unit of time of planting be given by the function $C(x, u)$. The total discounted cost of planting at the rate $x(t)$ in the period from $t = 0$ to $t = 5$, when the rate of interest is $r$, is then $\int_0^5 C(x, u)e^{-rt} dt$. Write down the necessary conditions for the problem

$$
\min \int_0^T C(x, \dot{x})e^{-rt} dt, \quad x(0) = 0, \quad x(5) \geq 1500
$$

(b) Solve the problem when $r = 0$, and $C(x, u) = g(u)$, with $g(0) = 0$, $g'(u) > 0$, and $g''(u) > 0$.

---

**CONTROL THEORY: BASIC TECHNIQUES**

A person who insists on understanding every tiny step before going to the next is liable to concentrate so much on looking at his feet that he fails to realize he is walking in the wrong direction.

—I. Stewart (1975)

Optimal control theory is a modern extension of the classical calculus of variations. Whereas the Euler equation dates back to 1744, the main result in optimal control theory, the maximum principle, was developed as recently as the 1950s by a group of Russian mathematicians (Pontryagin et al., 1962)). This principle gives necessary conditions for optimality in a wide range of dynamic optimization problems. It includes all the necessary conditions that emerge from the classical theory, but can also be applied to a significantly wider range of problems.

Since 1960, thousands of papers in economics have used optimal control theory. Its applications include, for instance, economic growth, inventory control, taxation, extraction of natural resources, irrigation, and the theory of regulation under asymmetric information.

This chapter contains some important results for the one state variable case that are used widely in the economics literature. ("What every young economist should know about optimal control theory.")

After the introductory Section 9.1, Section 9.2 deals with the simple case in which there are no restrictions on the control variable or on the terminal state. Although such problems can usually be solved by using the calculus of variations, their relative simplicity make them ideal as a starting point for introducing some of the main concepts and ideas in control theory.

Some additional concepts like the control region, piecewise continuous controls, and required regularity conditions are spelled out in Section 9.3.

Section 9.4 goes on to consider different alternative terminal conditions on the state variable. The brief Section 9.5 shows how to formulate a calculus of variations problem as an optimal control problem. The Euler equation is easily derived as an implication of the maximum principle.

Section 9.6 is concerned with sensitivity results: what happens to the optimal value function when the parameters change? In a growth theory setting, these parameters are: the beginning and end of the planning period, the initial capital stock; and the amount of capital to leave at the end of the planning period. It turns out that the adjoint variable can be given interesting economic interpretations, somewhat similar to those for Lagrange multipliers.
A rigorous proof of the maximum principle is beyond the level of this book. However, the Mangasarian sufficiency theorem is proved quite easily in Section 9.7. An important generalization (the Arrow sufficiency theorem) is also explained.

In some economic problems the final time (say the end of the planning period) is a variable to be chosen optimally along with the optimal control. Such problems are discussed in Section 9.8. Note 9.8.1 points out that the conditions in the Mangasarian or Arrow theorems are not sufficient for optimality in variable time period.

In optimal economic growth models there is often a discount factor in the objective function. Then a slight reformulation of the maximum principle, called the current value formulation, is frequently used. This approach is explained in Section 9.9.

Some economic models include a "scrap value" in the objective function. Section 9.10 explains how to adjust the maximum principle to take care of this case.

Most control models in the economic literature assume an infinite time horizon. Section 9.11 is an attempt to give some correct results and examples in this area where there is some confusion, even in leading textbooks.

Even control problems for which explicit solutions are unobtainable can sometimes be analysed by the phase diagram technique explained in Section 9.12.

### 9.1 The Basic Problem

Consider a system whose state at time \( t \) is characterized by a number \( x(t) \), the state variable. The process that causes \( x(t) \) to change can be controlled, at least partially, by a control function \( u(t) \). We assume that the rate of change of \( x(t) \) depends on \( t \), \( x(t) \), and \( u(t) \). The state at some initial point \( t_0 \) is typically known, \( x(t_0) = x_0 \). Hence, the evolution of \( x(t) \) is described by a controlled differential equation

\[
\dot{x}(t) = g(t, x(t), u(t)), \quad x(t_0) = x_0
\]

Suppose we choose some control function \( u(t) \) defined for \( t \geq t_0 \). Inserting this function into (1) gives a first-order differential equation for \( x(t) \). Because the initial point is fixed, a unique solution of (1) is usually obtained.

By choosing different control functions \( u(t) \), the system can be steered along different paths, not all of which are equally desirable. As usual in economic analysis, assume that it is possible to measure the benefits associated with each path. More specifically, assume that the benefits can be measured by means of the integral

\[
J = \int_{t_0}^{t} f(t, x(t), u(t)) \, dt
\]

where \( f \) is a given function. Here, \( J \) is called the objective or the criterion. Certain restrictions are often placed on the final state \( x(t_f) \). Moreover, the time \( t_f \) at which the process stops is not necessarily fixed. The fundamental problem that we study is:

Among all pairs \( (x(t), u(t)) \) that obey the differential equation in (1) with \( x(t_0) = x_0 \) and that satisfy the constraints imposed on \( x(t_f) \), find one that maximizes (2).

### Example 1

**Economic growth**

Consider the control problem

\[
\max_{u(t)} \int_{0}^{T} (1-s)f(k) \, ds, \quad \dot{k} = sf(k), \quad k(0) = k_0, \quad k(T) \geq k_T, \quad 0 \leq s \leq 1
\]

Here \( k = k(t) \) is the real capital stock of a country and \( f(k) \) is its production function. Moreover, \( s = s(t) \), the control variable, is the fraction of production set aside for investment, and it is natural to require that \( s \in [0, 1] \). The quantity \( (1-s)f(k) \) is the flow of consumption per unit of time. We wish to maximize the integral of this quantity over \([0, T] \), i.e., to maximize total consumption over the period \([0, T] \). The constant \( k_0 \) is the initial capital stock, and the condition \( k(T) \geq k_T \) means that we wish to leave a capital stock of at least \( k_T \) to those who live after time \( T \). (Example 9.6.3.b studies a special case of this model.)

### Example 2

**Oil extraction**

Let \( x(t) \) denote the amount of oil in a reservoir at time \( t \). Assume that at \( t = 0 \) the field contains \( K \) barrels of oil, so that \( x(0) = K \). If \( r(t) \) is the rate of extraction, then integrating each side of (4) yields \( x(t) - x(0) = - \int_{0}^{t} u(t) \, ds \), or \( x(t) = K - \int_{0}^{t} u(t) \, ds \) for each \( t \geq 0 \). That is, the amount of oil left at time \( t \) is equal to the initial amount \( K \), minus the total amount that has been extracted during the time span \([0, t] \), namely \( \int_{0}^{t} u(t) \, ds \). Differentiating gives

\[
\dot{x}(t) = -u(t), \quad x(0) = K
\]

Suppose that the market price of oil at time \( t \) is known to be \( p(t) \), so that the sales revenue per unit of time at \( t \) is \( q(t)u(t) \). Assume further that the cost \( C \) per unit of time depends on \( t, x, \) and \( u \), so that \( C = C(t, x, u) \). The instantaneous profit per unit of time at time \( t \) is then

\[
x(t, x(t), u(t)) = q(t)u(t) - C(t, x(t), u(t))
\]

If the discount rate is \( r \), the total discounted profit over the interval \([0, T] \) is

\[
\int_{0}^{T} \left[ q(t)u(t) - C(t, x(t), u(t)) \right] e^{-rt} \, dt
\]

It is natural to assume that \( u(t) \geq 0 \), and that \( x(T) \geq 0 \).

**Problem I** Find the rate of extraction \( u(t) \geq 0 \) that maximizes (++) subject to (s) and \( x(T) \geq 0 \) over a fixed extraction period \([0, T] \).

**Problem II** Find the rate of extraction \( u(t) \geq 0 \) and also the optimal terminal time \( T \) that maximizes (++) subject to (s) and \( x(T) \geq 0 \).

These are two instances of optimal control problems. Problem I has a fixed terminal time \( T \), whereas Problem II is referred to as a free terminal time problem. See Example 9.8.1.
9.2 A Simple Case

We begin by studying a control problem with no restrictions on the control variable and no restrictions on the terminal state—that is, no restrictions are imposed on the value of \( x(t) \) at \( t = t_1 \). Given the fixed times \( t_0 \) and \( t_1 \), our problem is

\[
\text{maximize} \quad \int_{t_0}^{t_1} f(t, x(t), u(t)) \, dt, \quad u(t) \in (-\infty, \infty)
\]

subject to

\[
\dot{x}(t) = g(t, x(t), u(t)), \quad x(t_0) = x_0, \quad x(t_1) \text{ fixed}, \quad x(t_1) \text{ free}
\]

Given any control function \( u(t) \) defined on \([t_0, t_1]\), the associated solution of the differential equation in (2) with \( x(t_0) = x_0 \) will usually be uniquely determined on the whole of \([t_0, t_1]\). A pair \((x(t), u(t))\) that satisfies (2) is called an admissible pair. Among all admissible pairs we search for an optimal pair, i.e., a pair of functions that maximizes the integral in (1).

Notice that the problem is to maximize the objective \( f(t, x(t), u(t)) \) subject to the constraint (2). Because this constraint is a differential equation on the interval \([t_0, t_1]\), it can be regarded as an infinite number of equality constraints, one for each time \( t \) in \([t_0, t_1]\).

Economists usually incorporate equality constraints in their optimization problems by forming a Lagrangian function, with a Lagrange multiplier corresponding to each constraint. Analogously, we associate a number \( p(t) \), called the costate variable, with the constraint (2) for each \( t \) in \([t_0, t_1]\). The resulting function \( p = p(t) \) is called the adjoint function associated with the differential equation. Corresponding to the Lagrangian function in the present problem is the Hamiltonian \( H \). For each time \( t \) in \([t_0, t_1]\) and each possible triple \((x, u, p)\), the Hamiltonian is defined by

\[
H(t, x, u, p) = f(t, x, u) + pg(t, x, u)
\]

A set of necessary conditions for optimality is given in the following theorem (some regularity conditions required are discussed in the next section):

**Theorem 9.2.1 (The Maximum Principle)**

Suppose that \((x^*(t), u^*(t))\) is an optimal pair for problem (1)–(2). Then there exists a continuous function \( p(t) \) such that, for each \( t \) in \([t_0, t_1]\),

\[
u = u^*(t) \text{ maximizes } H(t, x^*(t), u, p(t)) \text{ for } u \in (-\infty, \infty)
\]

\[
\dot{p}(t) = -H_p' (t, x^*(t), u^*(t), p(t)), \quad p(t_1) = 0
\]

**NOTE 1** The requirement that \( p(t_1) = 0 \) in (5) is called a transversality condition. So condition (5) tells us that in the case where \( x(t) \) is free, the adjoint variable vanishes at \( t_1 \).

The conditions in Theorem 9.2.1 are necessary, but not sufficient for optimality. The following theorem, which is a special case of Theorem 9.4.4 below, gives sufficient conditions:

1 The correspondence is rather loose. Something closer to the Lagrangian would be the function \( f(t, x, u) + pg(t, x, u) - k \).

If the requirement

\[
H(t, x, u, p(t)) \text{ is concave in } (x, u) \text{ for each } t \in [t_0, t_1]
\]

is added to the requirements in Theorem 9.2.1, then we obtain sufficient conditions. That is, if we find a triple \((x^*(t), u^*(t), p(t))\) that satisfies (2), (4), (5), and (6), then \((x^*(t), u^*(t))\) is optimal.

**NOTE 2** Changing \( u(t) \) on a small interval causes \( f(t, x, u) \) to change immediately. Moreover, at the end of this interval \( x(t) \) has changed and this change is transmitted throughout the remaining time interval. In order to steer the process optimally, the choice of \( u(t) \) at each instant of time must anticipate the future changes in \( x(t) \). In short, we have to plan ahead. In a certain sense, the adjoint function \( p(t) \) takes care of this need for forward planning. Equation (5) implies that \( p(t) = \int_t^{t_1} H_p'(t, x^*(s), u^*(s), p^*(s)) \, ds \).

**NOTE 3** If the problem is to minimize the objective in (1), then we can rewrite the problem as one of maximizing the negative of the original objective function. Alternatively, we could reformulate the maximum principle for the minimization problem: an optimal control will minimize the Hamiltonian, and convexity of \( H(t, x, u, p(t)) \) w.r.t. \((x, u)\) is the relevant sufficient condition.

Since the control region is \((-\infty, \infty)\), a necessary condition for (4) is that

\[
H_p'(t, x^*(t), u^*(t), p(t)) = 0
\]

If \( H(t, x, u, p(t)) \) is concave in \( u \), condition (7) is also sufficient for the maximum condition (4) to hold, because we recall that an interior stationary point for a concave function is (globally) optimal.

It is helpful to see how these conditions allow some simple examples to be solved.

**Example 1**

Solve the problem

\[
\text{max} \quad \int_0^T \left[ 1 - tx(t) - u^2(t) \right] \, dt, \quad \dot{x}(t) = u(t), \quad x(0) = x_0, \quad x(T) \text{ free}, \quad u \in \mathbb{R}
\]

where \( x_0 \) and \( T \) are given positive constants.

**Solution:** The Hamiltonian is \( H(t, x, u, p) = 1 - tx - u^2 + pu \), which is concave in \( u \), so the control \( u = u^*(t) \) maximizes \( H(t, x^*(t), u, p(t)) \) w.r.t. \( u \) if and only if it satisfies \( H_p' = -2u + p = 0 \). Thus \( u^*(t) = \frac{1}{2} p(t) \). Because \( H_p' = -t \), the conditions in (5) reduce to \( \dot{p}(t) = t \) and \( p(T) = 0 \). Integrating gives \( p(t) = \frac{1}{2} t^2 + C \) with \( \frac{1}{2} T^2 + C = 0 \), so

\[
p(t) = -\frac{1}{2} (T^2 - t^2)
\]

and then \( u^*(t) = -\frac{1}{2} (T^2 - t^2) \)
Because \( \dot{x}(t) = u^*(t) = -\frac{1}{2}(T^2 - t^2) \), integrating \( \dot{x}(t) = u^*(t) \) with \( x^*(0) = x_0 \) gives
\[
x^*(t) = x_0 - \frac{T^2}{2}t + \frac{1}{12}t^3
\]
Thus, there is only one pair \((x^*(t), u^*(t))\) that, together with \( p(t) \), satisfies both necessary conditions (4) and (5). We have therefore found the only possible path that can solve the problem. Because \( H(t, x, u, p) = -x^2 - cu^2 +pu \) is concave in \((x, u)\) for each fixed \(t\) (it is a sum of concave functions), \((x^*(t), u^*(t))\) is indeed optimal.

**EXAMPLE 2:** Solve the problem
\[
\max_{u(t) \in (-\infty, 0)} \int_0^T (-x^2 - \frac{1}{2}u^3)e^{-2t} dt, \quad \dot{x} = x + u, \quad x(0) = 1, \quad x(T) = x(T) \text{ free}
\]

**Solution:** The Hamiltonian is \( H(t, x, u, p) = -x^2 - \frac{1}{2}u^3e^{-2t} + p(x + u) \). The maximum principle states that if an admissible pair \((x^*(t), u^*(t))\) satisfies the problem, then there exists a function \( p \) defined on \([0, T]\) such that:
1. For every \( t \) in \([0, T]\), \( u = u^*(t) \) maximizes \( H(t, x^*(t), u, p(t)) \) for \( u \) in \((-\infty, \infty)\).
2. \( \dot{p}(t) = -H_x^t(t, x^*(t), u^*(t), p(t)) = 2x^*(t)e^{-2t} - p(t) \) and \( p(0) = 0 \).

Since \( H_x^t(t, x^*(t), u^*(t), p(t)) = 2x^*e^{-2t} + p \), it follows from (i) that \( u^*(t) = e^{2t}p(t) \). The equation \( \dot{x}^* = x^* + u^* \) then yields \( x^*(t) = x^*(0) + e^{2t}p(t) \). Thus \( x^* \) and \( p \) must satisfy the system
\[
\begin{align*}
\dot{x} &= x + e^{2t}p \\
\dot{p} &= 2e^{-2t}x - p
\end{align*}
\]
of two simultaneous equations in the plane, which appeared as Problem 6.5.3. The general solution is
\[
x = Ae^{2t+2x} + Be^{-2t+2x} \quad \text{and} \quad p = A\sqrt{2}e^{-x} + B\sqrt{2}e^{x}.
\]
It remains to determine the constants \( A \) and \( B \) so that \( x^*(0) = 1 \) and \( p(T) = 0 \). This yields \( A + B = 1 \) and \( A\sqrt{2}e^{-x} + B\sqrt{2}e^{x} = 0 \). The solution is
\[
A = \frac{1 + e^{2T}}{2e^{T}} - 1 \quad \text{and} \quad B = \frac{e^{-T}(1 + e^{2T})}{2e^{T}}.
\]
The corresponding functions \( x^*, u^* \), and \( p \) satisfy the conditions in the maximum principle, and \( H \) is concave w.r.t. \((x, u)\), so it follows from Theorem 9.2.2 that this is the optimal solution.

**EXAMPLE 3:** (A macroeconomic control problem) Consider once again the macroeconomic model of Example 8.2.2. If we drop the terminal constraint at the end of the planning period, we face the following control problem:
\[
\min_{u(t) \in (-\infty, 0)} \int_0^T [x(t)^2 + cu(t)^2] dt, \quad \dot{x}(t) = u(t), \quad x(0) = x_0, \quad x(T) \text{ free}
\]
where \( u(t) \in \mathbb{R} \) and \( c > 0 \), and \( T \) are given. Use the maximum principle to solve the problem.

**Solution:** We maximize \( \int_0^T [x(t)^2 + cu(t)^2] dt \). The Hamiltonian is \( H(t, x, u, p) = -x^2 - cu^2 + pu \), so \( H_x^t = -2x \) and \( H_u^t = -2cu + p \). A necessary condition for \( u = u^*(t) \) to maximize the Hamiltonian is that \( H_u^t = 0 \) at \( u = u^*(t) \), or that \( -2cu^* + p(t) = 0 \). Therefore, \( u^*(t) = p(t)/2c \). The differential equation for \( p(t) \) is
\[
\dot{p}(t) = -H_x^t(t, x(t), u^*(t), p(t)) = 2x^*(t)
\]
From \( \dot{x}^* = u^*(t) \) and \( u^*(t) = p(t)/2c \), we have
\[
\dot{x}^*(t) = \frac{p(t)}{2c}
\]
The two first-order differential equations (**) and (**) can be used to determine the functions \( p \) and \( x^* \). Differentiate (*) w.r.t. \( t \) and then use (**) to obtain \( \dot{p}(t) = 2x(t) = p(t)/c \), whose general solution is
\[
p(t) = Ae^{rt} + Be^{-rt}, \quad \text{where} \quad r = \frac{1}{\sqrt{c}}.
\]
Imposing the boundary conditions \( p(T) = 0 \) and \( p(0) = 2x^*(0) = 2x_0 \), implies that \( Ae^{rt} + Be^{-rt} = 0 \) and \( r(A - B) = 2x_0 \). These two equations determine \( A \) and \( B \), and yield \( A = 2x_0e^{-rt}/[r(e^{rt} + e^{-rt})] \) and \( B = -2x_0e^{rt}/[r(e^{rt} + e^{-rt})] \). Therefore,
\[
p(t) = \frac{2x_0e^{-rt} - e^{rt} - e^{-(T-t)}}{e^{rt} + e^{-rt}} \quad \text{and} \quad \dot{x}^*(t) = \frac{1}{2}p(t) = \frac{x_0(e^{rt} - e^{-(T-t)})}{e^{rt} + e^{-rt}}
\]
The Hamiltonian \( H = -x^2 - cu^2 + pu \) is concave in \((x, u)\), so by Mangasarian's theorem, this is the solution to the problem. (The same result was obtained in Example 8.5.2.)

**PROBLEMS FOR SECTION 9.2:**

Solve the control problems 1–5:

1. \[
\max_{u(t) \in (-\infty, 0)} \int_0^T [x(t) - u(t)]^2 dt, \quad \dot{x}(t) = -u(t), \quad x(0) = 0, \quad x(2) \text{ free}
\]
2. \[
\max_{u(t) \in (-\infty, 0)} \int_0^T [1 - u(t)]^2 dt, \quad \dot{x}(t) = x(t) + u(t), \quad x(0) = 1, \quad x(1) \text{ free}
\]
3. \[
\min_{u(t) \in (-\infty, 0)} \int_0^T [x(t) + u(t)]^2 dt, \quad \dot{x}(t) = -u(t), \quad x(0) = 0, \quad x(1) \text{ free}
\]
4. \[
\max_{u(t) \in (-\infty, 0)} \int_0^T [1 - 4x(t) - 2u(t)]^2 dt, \quad \dot{x}(t) = u(t), \quad x(0) = 0, \quad x(10) \text{ free}
\]
5. \[ \max_{x(0) \in (0,0,0)} \int_0^T (x - u^2) \, dt, \quad \dot{x} = x + u, \quad x(0) = 0, \quad x(T) \text{ free} \]

6. (a) Write down conditions (7) and (5) for the problem

\[ \max_{x(0) \in (0,0,0)} \int_0^T \left[ \phi(K) - c(I) \right] \, dt, \quad \dot{x} = I - \delta K, \quad K(0) = K_0, \quad K(T) \text{ free} \]

Here is an economic interpretation: \( K = K(t) \) denotes the capital stock of a firm, \( f(K) \) is the production function, \( q \) is the price per unit of output, \( I = I(t) \) is investment, \( c(I) \) is the cost of investment, \( \delta \) is the rate of depreciation of capital, \( K_0 \) is the initial capital stock, and \( T \) is the fixed planning horizon.

(b) Let \( f(K) = K - 0.05K^2, \quad q = 1, \quad c(I) = I^2, \quad \delta = 0.1, \quad K_0 = 10, \quad T = 10 \). Derive a second-order differential equation for \( K \), and explain how to find the solution.

9.3 Regularity Conditions

In most applications of control theory to economics, the control functions are explicitly or implicitly restricted in various ways. For instance, \( u(t) \geq 0 \) was a natural restriction in the oil extraction problem of Section 9.1; it means that one cannot pump oil back into the reservoir.

In general, assume that \( u(t) \) takes values in a fixed subset \( U \) of the reals, called the control region. In the oil extraction problem, then, \( U = [0, \infty) \). Actually, an important aspect of control theory is that the control region may be a closed set, so that \( u(t) \) can take values at the boundary of \( U \). (In the classical calculus of variation, by contrast, one usually considered open control regions, although developments around 1930–1940 paved the way for the modern theory.)

What regularity conditions is it natural to impose on the control function \( u(t) \)? Among the many papers in economics literature that use control theory, the majority assume implicitly or explicitly that the control functions are continuous. Consequently, many of our examples and problems will deal with continuous controls. Yet in some applications, continuity is too restrictive. For example, the control variable \( u(t) \) could be the fraction of investment allocated to one plant, with the remaining fraction \( 1 - u(t) \) is allocated to another. Then it is natural to allow control functions that suddenly switch all the investment from one plant to the other. Because they alternate between extremes, such functions are often called bang-bang controls. A simple example of such a control is

\[ u(t) = \begin{cases} 1 & \text{for } t \in [0, t'] \\ 0 & \text{for } t \in (t', t] \end{cases} \]

which involves a single shift at time \( t' \). In this case \( u(t) \) is piecewise continuous, with a jump discontinuity at \( t = t' \).

By definition, a function of one variable has a finite jump at a point of discontinuity if it has (finite) one-sided limits from both above and below at that point. A function is piecewise continuous if it has at most a finite number of discontinuities in each finite interval, with finite jumps at each point of discontinuity. (The value of a control \( u(t) \) at each isolated point of discontinuity will affect neither the integral objective nor the state, but let us agree to choose the value of \( u(t) \) at a point of discontinuity \( t' \) as the left-hand limit of \( u(t) \) at \( t' \). Then \( u(t) \) will be left-continuous, as illustrated in Fig. 1.) Moreover, if the control problem concerns the time interval \([0, t_1]\), we shall assume that \( u(t) \) is continuous at both end points of this interval.

What is meant by a "solution" of \( \dot{x} = g(t, x, u) \) when \( u = u(t) \) has discontinuities? A solution is a continuous function \( x(t) \) that has a derivative that satisfies the equation, except at points where \( u(t) \) is discontinuous. The graph of \( x(t) \) will, in general, have "kinks" at the points of discontinuity of \( u(t) \), and it will usually not be differentiable at these kinks. It is, however, still continuous at the kinks.

For the oil extraction example in Problem 9.1.2, Fig. 1 shows one possible control function, whereas Fig. 2 shows the corresponding development of the state variable. The rate of extraction is initially a constant \( u_0 \) on the interval \([0, t']\), then a different constant \( u_1 \) (with \( u_1 < u_0 \)) on \([t', t'']\). Finally, on \([t'', T]\), the rate of extraction \( u(t) \) gradually declines from a level lower than \( u_1 \) until the field is exhausted at time \( T \). Observe that the graph of \( x(t) \) is connected, but has kinks at \( t' \) and \( t'' \).

![Figure 1](image1)

![Figure 2](image2)

So far no restrictions have been placed on the functions \( g(t, x, u) \) and \( f(t, x, u) \). For the analysis presented in this chapter, it suffices to assume that \( f, g, \) and their first-order partial derivatives w.r.t. \( x \) and \( u \) are continuous in \((t, x, u)\). These continuity properties will be implicitly assumed from now on.

Necessary Conditions, Sufficient Conditions, and Existence

In static optimization theory there are three main types of result that can be used to find possible global solutions: theorems giving necessary conditions for optimality (typically, first-order conditions); theorems giving sufficient conditions (typically, first-order conditions supplemented by appropriate concavity/convexity requirements); and existence theorems (typically, the extreme value theorem).

In control theory the situation is similar. The maximum principle, in different versions, gives necessary conditions for optimality, i.e. conditions that an optimal control must satisfy. These conditions do not guarantee that the maximization problem has a solution.
4. $A = K(e^t + (r - \rho)/br + L$. The constants $K$ and $L$ are determined by $A(0) = A_0$, $A(T) = A_T$.
5. (a) The conditions are $-\left(d/dt\right)C(t, x)e^{-\alpha t} = 0$ and $C(5, x) \geq 0 \iff (t, x) > 1500$.
   (b) $x(t) = 3000$, implying that planting takes place at the constant rate of 300 hectares per year.

Chapter 9

9.2
1. The Hamiltonian is $H = e^t e - u^2 + p(u)$, so $H_x = e'$ and $H_u = -2a - p$. Because $U = (\infty, \infty, 0)$, (7) implies that $u = \frac{1}{2}p$. (5) reduces to $\dot{p} = -e^u$, $p(t) = e^u$, from which it follows that $p(t) = e^u$. From

$$u(t) = \begin{cases} \frac{1}{2}p(t) & \text{if } t < [0, t^*] \\ \frac{1}{2}(3 + \sqrt{12}) & \text{if } t \in [t^*, 2] \\ \frac{1}{2}(3 + \sqrt{12}) & \text{if } t \in [t^*, 2] \end{cases}$$

9. (a) Show that $p_u + p(t)$ must be 0. Hence, $p(t) = 0$, and $H = 0$.
   (b) $\max \int_0^t u(t) dt = \max \int_0^t u(t) dt = \max \int_0^t e^u dt = \frac{1}{2}(3 + \sqrt{12})$

10. (a) Obviously, $u(t) = 0$ is the only admissible control, so it is optimal.
   (b) With $W = -\rho^2 + pu^2$, the Hamiltonian control is $H = -\rho^2 + pu^2$, which minimizes $u(t) = 0$.

9.5
1. The Euler equation is $\dot{x} = -e^u$. The solution is $x = -e^{u^2} + e^{u^2} + 1$. To solve it as a control problem, put $x = u$.

2. $x(t) = A^2(1 - 2e^{t^2})$, with $A = (2e^{t^2} - 1)$.

3. The Euler equation is $\dot{x} = -\frac{1}{2}\left(-2 - 2e^{-2u^2}\right)$, which implies $\dot{x} = A^2t^2 - 1$, for some constant $A$. Integrate and use the boundary conditions to get the solution $x = e^{-u^2} + 1$. (Alternative form of the Euler equation: $\dot{x} = -\frac{1}{2}\left(-2 - 2e^{-2u^2}\right)$, $x(t) = 1 - t$.

4. The Euler equation reduces to $\frac{d}{dt}\left(\dot{x}^2 - 2\dot{x}e^{-u^2}\right) = -2\ddot{x}e^{-u^2}$, so $\dot{x} = e^{-u^2}$ and $x(t) = 1 - t$.

5. (a) With $H = (x - b)^2 - g(x) + px$, the conditions are: (i) $x$ maximizes $U(x) - k(x) - b^2 + p(x)$ for $x \geq 0$; (ii) $\dot{x} = p$, $p(t) = 0$. From (i) we immediately get $p(t) = g(t) - 1$. Moreover, $H = (x - b)^2 - g(x)$ implies that $x(t) = 1 - t$.

9.6
1. (a) $u^*(t) = T - t, x^*(t) = x_T + T - \frac{1}{2}t^2$. (b) $x(0, T) = x_T + T^2$ and the relevant equalities in (2) are easily verified.

2. $H^*(T) = x^*(T) + p(T)x^*(T) = x(T) + T$, so using the expression for $V(t)$, $V(T) = H^*(T)$, it follows that $V(T) = H^*(T)$.

3. Using the results in the answers to Problem 9.4.4 (b), we find: $V(0, x) = x_T + x_T = 0$. Then

$$V^*(T) = x_T - T – \frac{1}{2}T^2,$$

for $T > \frac{1}{2}T^2$.

4. (a) For $T \leq \frac{1}{2}T^2$, we have $u(t) = 0$ and $x(t) = 1$, with $p(t) = 2e^{-t^2/2}(1 - e^{-t^2/2})$.

(b) $V_T = 1 - \frac{1}{2}T^2$, $x(0, T) = x_T + \frac{1}{2}T^2 - \frac{1}{2}T^2$,

5. (a) $x(t) = x_T + \frac{1}{2}t^2$, $u(t) = 0$, and $x(t) = 1$, with $\dot{x} = e^{-u^2}$.

(b) $x(0, T) = x_T + \frac{1}{2}T^2$, $x_T = 1$, and $x = e^{-u^2}$.

6. $u(t) = 0$ and $V_T = 0$. Then $V_T = x_T - T – \frac{1}{2}T^2$.

7. (a) $u(t) = 0$ and $V_T = x_T - T – \frac{1}{2}T^2$. (b) $u(t) = 0$ and $V_T = x_T - T – \frac{1}{2}T^2$.
sets $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$, and, as $n$ increases, $A_n$ will "tend to" the set $A = \{(x, y) : 0 < x^2 + y^2 \leq 1\}$, because $\bigcap_{n=1}^{\infty} A_n = A$. The double integral of $f$ is defined over the set $A_n$ for each $n = 1, 2, \ldots$. Indeed, by introducing polar coordinates, we obtain

$$I_n = \int_{A_n} \frac{1}{(x^2 + y^2)} \, dx \, dy = \int_0^{2\pi} \left( \int_{r_1}^{r_2} \frac{1}{r^2} \, r \, dr \right) \, d\theta = \int_0^{2\pi} \left( \int_{1/n}^{\infty} \frac{1}{r} \, dr \right) \, d\theta$$

It follows that $I_n = 2\pi \ln n$ for $p = 1$, while $I_n = \pi(1 - n^{2(p-1)})/(1 - p)$ for $p \neq 1$. If $p < 1$, then $n^{2(p-1)} \to 0$ as $n \to \infty$, and $I_n \to \pi / (1 - p)$. If $p \geq 1$, we see that $I_n$ does not tend to any limit. On the basis of this observation, if $0 < p < 1$, we say that the integral is convergent, with a value $\pi / (1 - p)$, whereas we say that it is divergent if $p \geq 1$.

---

5. DIFFERENTIAL EQUATIONS I:

FIRST-ORDER EQUATIONS IN ONE VARIABLE

...the task of the theory of ordinary differential equations is to reconstruct the past and predict the future of the process from a knowledge of this local law of evolution.

—V.I. Arnold (1973)

---

Economists often study the changes over time in economic variables like national income, the interest rate, the money supply, oil production, or the price of wheat. The laws of motion governing these variables are usually expressed in terms of one or more equations.

If time is regarded as continuous and the equations involve unknown functions and their derivatives, we find ourselves considering differential equations. In macroeconomic theory especially, but also in many other areas of economics, a certain knowledge of differential equations is essential. Another example is finance theory, where the pricing of options now requires quite advanced methods in the theory of differential equations.

The systematic study of differential equations was initiated by Newton and Leibniz in the seventeenth century, and this topic is still one of the most important in mathematics.

After the introductory Section 5.1, the short Section 5.2 merely points out how to draw a direction diagram, and how solving a differential equation is equivalent to finding a curve whose tangent at each point is given by the direction diagram.

Section 5.3 gives a systematic discussion of separable differential equations, i.e. equations of the form $x = f(yg(x))$.

Section 5.4 concentrates on the special properties of first-order linear differential equations, first with constant and then with variable coefficients. Several economic examples are studied.

Section 5.5 deals with exact equations and integrating factors. Exact equations appear less frequently in economics and this section is therefore in small print.

Although only a few types of differential equations have solutions given by explicit formulas, Section 5.6 shows how a clever choice of new variables can sometimes help with seemingly insoluble equations.

Section 5.7 considers what qualitative properties of a solution can be inferred, even if the equation cannot be solved analytically.

Finally, Section 5.8 is concerned with existence and uniqueness theorems for first-order equations.
5.1 Introduction

What is a differential equation? As the name suggests, it is an equation. Unlike ordinary algebraic equations, in a differential equation:

(A) The unknown is a function, not a number.
(B) The equation includes one or more of the derivatives of the function.

An ordinary differential equation is one for which the unknown is a function of only one variable. Partial differential equations are equations where the unknown is a function of two or more variables, and one or more of the partial derivatives of the function are included.

In this chapter we restrict attention to first-order (ordinary) differential equations—that is, equations where only the first-order derivatives of the unknown functions of one variable are included. Three typical examples are:

\[ \dot{x} = ax, \quad \dot{x} + ax = b, \quad \dot{x} + ax = bx^2 \]

With suitably chosen constants, these describe natural growth, growth towards a limit, and logistic growth, respectively. (Recall that we often use dot notation for the derivative, \( \dot{x} = dx/dt \), especially when the independent variable is time \( t \).) Other examples of first-order differential equations are

(a) \( \dot{x} = x + t \)  
(b) \( \dot{K} = \alpha K + H_0 e^{\mu t} \)  
(c) \( \dot{k} = s f(k) - \lambda k \)

In Examples 5.4.3 and 5.7.3, respectively, we shall give equations (b) and (c) interesting economic interpretations, both concerning the evolution of an economy's capital stock.

Solving equation (a), for instance, means finding all functions \( x(t) \) such that, for every value of \( t \), the derivative \( x'(t) \) of \( x(t) \) is equal to \( x(t) + t \). In equation (b), \( K(t) \) is the unknown function, whereas \( \alpha, \sigma, H_0, \) and \( \mu \) are constants. In equation (c), \( f(k) \) is a given function, whereas \( s \) and \( \lambda \) are constants. The unknown function is \( k = k(t) \).

**NOTE** We often use \( t \) to denote the independent variable. This is because most differential equations that appear in economics have time as their independent variable. The following theory is valid even if the independent variable is not time, however.

A first-order differential equation is written

\[ \dot{x} = F(t, x) \]

where \( F \) is a given function of two variables and \( x = x(t) \) is the unknown function. A solution of (1) in an interval \( I \) of the real line is any differentiable function \( \phi \) defined on \( I \) such that \( x = \phi(t) \) satisfies (1), that is \( \phi(t) = F(t, \phi(t)) \) for all \( t \) in \( I \). The graph of a solution is called a solution curve or an integral curve.

The equations (a), (b), and (c) are all of the form (1). For example, (a) becomes \( dx/dt = F(t, x) = x + t \).

1 Usually we assume that the interval \( I \) is open, but sometimes it is useful to allow closed (or half-open) intervals. If \( I \) is a closed interval, a solution is required to be continuous on \( I \) and to satisfy (1) in the interior of \( I \).

**Example 1**

Consider the differential equation

\[ \dot{x} = x + t \]

(a) Show that both \( x = -t - 1 \) and \( x = e^t - t - 1 \) are particular solutions of the equation over the entire real line.

(b) More generally, show that \( x = Ce^t - t - 1 \) is a solution of (1) for all \( t \), whatever the choice of the constant \( C \).

(c) Show that \( x = e^t - 1 \) is not a solution of (1).

**Solution:**

(a) If \( x = -t - 1 \), then \( \dot{x} = -1 \) and \( x + t = (-t - 1) + t = -1 \). Hence, \( \dot{x} = x + t \) for all \( t \) in this case. If \( x = e^t - t - 1 \), then \( \dot{x} = e^t - 1 \) and \( x + t = (e^t - t - 1) + t = e^t - 1 \). Again we see that (a) is satisfied for all \( t \).

(b) When \( x = Ce^t - t - 1 \), we have \( \dot{x} = Ce^t - 1 = x + t \) for all \( t \).

(c) If \( x = e^t - 1 \), then \( \dot{x} = e^t \) and \( x + t = e^t + t - 1 \). In this case, \( \dot{x} \) is only equal to \( x + t \) for \( t = 1 \), so \( x = e^t - 1 \) is not a solution of equation (1) on any interval.

Example 1 provides the fact that a differential equation usually has infinitely many solutions. We found that \( x = Ce^t - t - 1 \) was a solution of \( \dot{x} = x + t \) for each choice of the constant \( C \). The answer to Problem 5.4.3 shows that no other function satisfies the equation.

The set of all solutions of a differential equation is called its general solution, while any specific function that satisfies the equation is called a particular solution.

A first-order differential equation usually has a general solution that depends on one constant. (Problem 5.5 shows why we must use the word "usually" in this statement.) If we require the solution to pass through a given point in the \( x \)-plane, then the constant is determined uniquely, except in special cases.

**Example 2**

Assuming that the general solution is \( x(t) = Ce^t - t - 1 \), find the solution of \( \dot{x} = x + t \) that passes through the point \( (t, x) = (0, 1) \).

**Solution:** To make the solution \( x(t) = Ce^t - t - 1 \) pass through \((t, x) = (0, 1)\), we must have \( x(0) = 1 \). Hence, \( 1 = Ce^0 - 0 - 1 \), implying that \( C = 2 \). The required solution, therefore, is \( x(t) = 2e^t - t - 1 \).

The problem in Example 2 is this: Find the unique function \( x(t) \) such that

\[ \dot{x}(t) = x(t) + t \quad \text{and} \quad x(0) = 1 \]

\[ (*) \]

If \( t = 0 \) denotes the initial time, then \( x(0) = 1 \) is called an initial condition and we call (*) an initial-value problem.

Such initial-value problems arise naturally in many economic models. For instance, suppose an economic growth model involves a first-order differential equation for the accumulation of capital over time. The initial stock of capital is historically given, and therefore helps to determine the unique solution of the equation.
Qualitative Theory

When the theory of differential equations was first developed, mathematicians primarily tried to find explicit solutions for some special types of equations. It became increasingly obvious, however, that only very few equations could be solved this way. In many cases, moreover, explicit formulas for the solutions are not really needed. Instead, the main interest is in a few important properties of the solution. As a result, the theory of differential equations includes many results concerning the general behaviour of the solutions. This is the so-called qualitative theory. Its main results include existence and uniqueness theorems, sensitivity analysis, and investigations of the stability of equilibria. Such topics are of both theoretical interest and practical importance, and will be discussed in some detail.

Along with this qualitative theory, much work has been put into developing useful numerical methods for finding approximate solutions of differential equations. Computers are playing an increasingly important role here, but these developments are not discussed here.

PROBLEMS FOR SECTION 5.1

1. Show that \( x(t) = Ce^{-t} + \frac{1}{2}t \) is a solution of the differential equation \( \dot{x}(t) + x(t) = c \) for all values of the constant \( C \).

2. Show that \( x = C e^t \) is a solution of the differential equation \( \dot{x} = 2t \) for all choices of the constant \( C \). Find in particular the integral curve through \((1, 2)\).

3. Show that any function \( x = x(t) \) that satisfies the equation \( x^2 + x + C \) is a solution of the differential equation \( (1 + t)x + \dot{x} = -x^2 \). (Hint: Differentiate \( x^2 + C \) implicitly with \( t \).)

4. In each of the following cases, show that any function \( x = x(t) \) that satisfies the equation on the left is a solution of the corresponding differential equation on the right.
   (a) \( x^2 = 2at \), \( 2x = 2x^2 + a \) (\( a \) is a constant)
   (b) \( \frac{dx}{dt} + e^{x(t)}(x + 1) + C = 0 \), \( x(t) = e^{x(t)^2} 
   (c) \( (1 - t)x^2 = t^2 \), \( 2x^2 = x(x^2 + 3t^2) \)

5. Show that \( x = Ct - C^2 \) is a solution of the differential equation \( \dot{x} = x - t \), for all values of the constant \( C \). Then show that it is not the general solution because \( x = \frac{1}{2}t^2 \) is also a solution.

HARDER PROBLEMS

6. The function \( x = x(t) \) satisfies \( x(0) = 0 \) and the differential equation \( \dot{x} = (1 + x^2)t \), for all \( t \). Prove that \( t = 0 \) is a global minimum point for \( x(t) \), and that the function \( x(t) \) is convex for all \( t \). (Hint: You do not have to solve the equation.)

SECTION 5.2 / THE DIRECTION IS GIVEN: FIND THE PATH!

5.2 The Direction is Given: Find the Path!

Consider again the differential equation \( \dot{x} = x + t \), which was studied in Examples 5.1.1 and 5.1.2. If \( x(t) \) is a solution, then the slope of the tangent to the graph (or integral curve) at the point \((t, x)\) is equal to \( x + t \). At the point \((0, 0)\) the slope is therefore equal to 0, whereas at \((1, 2)\) the slope is 3, and so on. In Fig. 1, we have drawn small straight-line segments with slopes \( x + t \) through several points in the \( x-t \)-plane. This gives us a so-called direction diagram (or slope field) for the differential equation \( \dot{x} = x + t \). If an integral curve passes through one of these points, it will have the corresponding line segment as its tangent. This allows us to sketch curves that follow the direction of the line segments, and get a general impression of what the integral curves of \( \dot{x} = x + t \) must look like.

A direction diagram like this can be drawn for any differential equation of the form \( \dot{x} = F(t, x) \). (Computer programs like Maple and Mathematica enable us to draw direction diagrams and solution curves with ease.) Whether or not it is possible to solve the equation explicitly, a direction diagram can give a rough but useful indication of how the integral curves behave. In a nutshell, the problem of solving the differential equation \( \dot{x} = F(t, x) \) can be put like this: the direction is given, find the path!

PROBLEMS FOR SECTION 5.2

1. Draw a direction diagram for the differential equation \( \dot{x} = x/2 \) and draw some integral curves.

2. Draw a direction diagram for the differential equation \( \dot{x} = -t/x \) and draw the integral curve through \((0, 2)\).
5.3 Separable Equations

Suppose that \( \dot{x} = F(t, x) \), where \( F(t, x) \) can be written as a product of two functions, one of which depends only on \( t \) and the other only on \( x \). Specifically, suppose that

\[
\dot{x} = f(t)g(x)
\]

We say that this differential equation is separable. For instance, \( \dot{x} = -2tx^2 \) is obviously separable, whereas \( \dot{x} = t^2 + x \) is not. (Problem 7 offers practice in deciding if a given equation is separable or not. Since separable equations are among those that can be solved in terms of integrals of known functions, it is useful to learn to distinguish between separable and nonseparable equations.)

A particular solution of (1) arises if \( g(x) \) has a zero at \( x = a \), so that \( g(a) = 0 \). In this case \( x(t) \equiv a \) will be a solution of the equation, because the right-hand sides are both zero for all \( t \). For instance, \( \dot{x} = (x+1)(x-3) \) has the two particular solutions \( x(t) \equiv -1 \) and \( x(t) \equiv 3 \). (In addition, \( x = -1 + 6/(1 - Ce^{at}) \) is a solution for all values of the constant \( C \). See Example 4 with \( B = 1, a = -1 \), and \( b = 3 \).)

Using differential notation, a general method for solving (1) can be expressed as follows:

**METHOD FOR SOLVING SEPARABLE DIFFERENTIAL EQUATIONS:**

(A) Write equation (1) as

\[
\frac{dx}{dt} = f(t)g(x)
\]

(B) Separate the variables:

\[
\frac{dx}{g(x)} = f(t) \, dt
\]

(C) Integrate each side:

\[
\int \frac{dx}{g(x)} = \int f(t) \, dt
\]

(D) Evaluate the two integrals (if possible) to obtain a solution of (e) (possibly in implicit form). Solve for \( x \), if possible.

(E) In addition, every zero \( x = a \) of \( g(x) \) gives the constant solution \( x(t) \equiv a \). To justify the method, suppose that \( x = \psi(t) \) is a function defined in an interval \( I \) such that \( g(\psi(t)) \neq 0 \) throughout \( I \). Then \( x = \psi(t) \) will solve (1) iff

\[
\int \frac{\psi'(t)}{g(\psi(t))} \, dt = \int f(t) \, dt
\]

for all \( t \) in \( I \). But these two functions are equal in \( I \) iff

\[
\int \frac{\psi(t)}{g(\psi(t))} \, dt = \int f(t) \, dt
\]

SECTION 5.3 / SEPARABLE EQUATIONS

Suppose we substitute \( x = \psi(t) \), so that \( dx = \psi'(t) \, dt \) in the integral on the left-hand side. Then according to the rule of integration by substitution, the last equation is equivalent to

\[
\int \frac{dx}{g(x)} = \int f(t) \, dt
\]

That, \( G(x) = F(t) + C \), where \( G'(x) = 1/g(x) \), \( F(t) = f(t) \), and \( C \) is a constant.

**NOTE 1** Suppose the function \( G(x) \) is defined on an interval \( I \) where either \( g(x) > 0 \) everywhere, or \( g(x) < 0 \) everywhere. If \( G(I) = R \), then a solution \( x(t) \) exists for all \( t \in \mathbb{R} \), with values in \( I \). But if \( G(I) \) is a proper subset of \( \mathbb{R} \), then \( x(t) = G^{-1}(F(t) + C) \) is a solution only for a restricted range of values of \( C \) and \( t \).

**EXAMPLE 1** Solve the differential equation

\[
\frac{dx}{dt} = -2tx^2
\]

and find the integral curve that passes through \( (t, x) = (1, -1) \).

**Solution:** We observe first that \( x(t) \equiv 0 \) is one (trivial) solution. But this does not go through \( (1, -1) \), so we follow the recipe:

Separate:

\[
\frac{dx}{x^2} = 2t \, dt
\]

Integrate:

\[
-x^{-1} = \int 2t \, dt
\]

Evaluate:

\[
\frac{1}{x} = t^2 + C
\]

It follows that the general solution is

\[
x = \frac{1}{t^2 + C}
\]

To find the integral curve that passes through \((1, -1)\), set \( x = -1 \) and \( t = 1 \) to get \( C = -2 \). The graph is shown in Figure 1.

**Figure 1** The solution curves \( x = 1/(t^2 + C) \) for particular values of \( C \).
EXAMPLE 2  Solve the differential equation \( \frac{dx}{dt} = \frac{t^3}{x^5 + 1} \).

Solution: We use the previous method, with \( f(t) = t^3 \) and \( g(x) = 1/(x^5 + 1) \). Because \( g(x) \) is never zero, there are no constant solutions. We proceed as follows:

Separate:
\[
(x^5 + 1) \, dx = t^3 \, dt.
\]
Integrate:
\[
\int (x^5 + 1) \, dx = \int t^3 \, dt.
\]
Evaluate:
\[
\frac{1}{5} x^6 + x = \frac{1}{4} t^4 + C.
\]

The desired functions \( x = x(t) \) are those that satisfy the last equation for all \( t \).

NOTE 2  We usually say that we have solved a differential equation even if the unknown function (as shown in Example 2) cannot be expressed explicitly. The important point is that we have found an equation involving the unknown function where the derivative of that function does not appear.

EXAMPLE 3  (Economic growth) Let \( X = X(t) \) denote the national product, \( K = K(t) \) the capital stock, and \( L = L(t) \) the number of workers in a country at time \( t \). Suppose that, for all \( t \geq 0 \),

(a) \( X = AK^{1-\alpha}L^\alpha \),
(b) \( \dot{K} = sX \),
(c) \( L = L_0 e^{\lambda t} \),

where \( A, \alpha, s, L_0, \) and \( \lambda \) are all positive constants, with \( \alpha < 1 \). Derive from these equations a single differential equation to determine \( K = K(t) \), and find the solution when \( K(0) = K_0 > 0 \). (This is a special case of the Solow–Swan model discussed in Example 5.3.1. In (a) we have a Cobb–Douglas production function, (b) says that aggregate investment is proportional to output, whereas (c) implies that the labour force grows exponentially.)

Solution: From the equations (a)–(c), we derive the single differential equation

\[
\dot{K} = \frac{dK}{dt} = sAK^{1-\alpha}L^\alpha = sAL_0^\alpha e^{\lambda t} K^{1-\alpha}.
\]

This is clearly separable. Using the recipe yields:

\[
K^{1-\alpha} \, dK = sAL_0^\alpha e^{\lambda t} \, dt,
\]

\[
\int K^{1-\alpha} \, dK = \int sAL_0^\alpha e^{\lambda t} \, dt,
\]

\[
\frac{1}{\alpha} K^\alpha = \frac{1}{\alpha} sAL_0^\alpha e^{\lambda t} + C
\]

If we put \( C_1 = aC \), we get

\[
K^\alpha = (sA/\alpha)L_0^\alpha e^{\lambda t} + C_1.
\]

If \( K = K_0 \) for \( t = 0 \), we get

\[
C_1 = K_0^\alpha - (sA/\alpha)L_0^\alpha.
\]

Therefore the solution is

\[
K = \left[ K_0^\alpha + (sA/\alpha)L_0^\alpha (e^{\lambda t} - 1) \right]^{1/\alpha}.
\]

See Problem 9 for a closer examination of this model.  

EXAMPLE 4  Solve the following differential equation when \( a \neq b \):

\[
\frac{dx}{dt} = B(x-a)(x-b).
\]

In particular, find the solution when \( B = -1/2, a = -1, \) and \( b = 2 \), and draw some integral curves in this case.

Solution: Observe that both \( x = a \) and \( x = b \) are trivial solutions of the equation. In order to find the other solutions, separate the variables as follows. First, put all terms involving \( x \) on the left-hand side, and all terms involving \( t \) on the right-hand side. Then integrate, to get

\[
\int \frac{1}{(x-a)(x-b)} \, dx = \int B \, dt.
\]

The next step is to transform the integrand on the left. We find that

\[
\frac{1}{(x-a)(x-b)} = \frac{1}{b-a} \left( \frac{1}{x-b} - \frac{1}{x-a} \right).
\]

(Verify this by expanding the right-hand side.) Hence,

\[
\int \frac{1}{(x-a)(x-b)} \, dx = \frac{1}{b-a} \left( \ln |x-b| - \ln |x-a| \right) = \frac{1}{b-a} \ln |x-b| - \ln |x-a|.
\]

Except for an additive constant, the last expression equals

\[
\frac{1}{b-a} \ln \frac{|x-b|}{|x-a|} = \frac{1}{b-a} \ln \frac{|x-b|}{|x-a|}.
\]

So, for some constant \( C_1 \), the solution is

\[
\frac{1}{b-a} \ln \frac{|x-b|}{|x-a|} = Bt + C_1 \quad \text{or} \quad \ln \frac{|x-b|}{|x-a|} = B(b-a)t + C_2
\]

with \( C_2 = C_1(b-a) \). So

\[
\frac{x-b}{x-a} = e^{B(b-a)t+C_2} \quad \text{or} \quad \frac{x-b}{x-a} = C e^{B(b-a)t}
\]

after defining the new constant \( C = C e^{C_1(b-a)} \). Solving this last equation for \( x \) finally gives

\[
\frac{dx}{dt} = B(x-a)(x-b) \iff x = \frac{b-a}{1-Ce^{B(b-a)t}} \quad \text{or} \quad \frac{dx}{dt} = B(x-a)(x-b)
\]

For \( B = -1/2, a = -1, \) and \( b = 2 \), the differential equation is

\[
\frac{dx}{dt} = -\frac{1}{2}(x+1)(x-2).
\]

Note that \( \lambda \) is positive for \( x > -1 \) and \( 2 \). Hence, the integral curves rise with \( t \) in the horizontal strip between lines \( x = -1 \) and \( x = 2 \). In the same way, we can see directly from the differential equation that the integral curves are falling above and below this strip. In addition to the constant solutions \( x = -1 \) and \( x = 2 \), indicated by dashed horizontal lines in Fig. 2, we see that the general solution of the differential equation \( \dot{x} = -\frac{1}{2}(x+1)(x-2) \) is

\[
x = -1 + \frac{3}{1-Ce^{-B(b-a)t}}
\]

Some of the associated integral curves are shown in Fig. 2.
Problems for Section 5.3

1. Solve the equation \( x' \cdot x = t + 1 \). Find the integral curve through \((t, x) = (1, 1)\).

2. Solve the following differential equations:
   \[
   \begin{align*}
   (a) \quad \dot{x} &= t^2 - t \\
   (b) \quad \dot{x} &= t e^{t^2} - t \\
   (c) \quad e^t \dot{x} &= t + 1
   \end{align*}
   \]

3. Find the general solutions of the following differential equations. Also find the integral curves through the indicated points:
   \[
   \begin{align*}
   (a) \quad t \dot{x} &= x(1 - t), \quad (t_0, x_0) = (1, 1/e) \\
   (b) \quad (1 + t^2) \dot{x} &= t^2 x, \quad (t_0, x_0) = (0, 2) \\
   (c) \quad \dot{x} &= t, \quad (t_0, x_0) = (\sqrt{2}, 1) \quad (d) \quad e^{\dot{x}} - x^2 - 2e^x = 1, \quad (t_0, x_0) = (0, 0)
   \end{align*}
   \]

4. Find the general solution of \( \dot{x} + a(t)x = 0 \). In particular, when \( a(t) = a + b e^{ct} \) (\( a, b, c \) are positive; \( c \neq 1 \)) show that the solution of the equation can be written in the form \( x = C e^{pt} e^{qt} \), where \( p \) and \( q \) are constants determined by \( a, b, c \), and \( C \), whereas \( C \) is an arbitrary constant. (This is Gompertz-Makeham's law of mortality.)

5. Explain why biological populations that develop as suggested in the figures A and B below cannot be described by differential equations of the form \( N' = f(N) \), no matter how the function \( f \) is chosen. \( (N(t)) \) is the size of the population at time \( t \).

6. Find \( x = x(t) \) when \( E_i x = \frac{\dot{x}}{x} \), the elasticity of \( x(t) \) w.r.t. \( t \), satisfies the following equations for all \( t \):
   \[
   \begin{align*}
   (a) \quad E_i x &= a \\
   (b) \quad E_i x &= a t + b \\
   (c) \quad E_i x &= a x + b
   \end{align*}
   \]

7. Decide which of the following differential equations are separable:
   \[
   \begin{align*}
   (a) \quad \dot{x} &= x^3 - 1 \\
   (b) \quad \dot{x} &= x t + t \\
   (c) \quad \dot{x} &= x t + t^2 \\
   (d) \quad \dot{x} &= e^{\sqrt{1 + x^2}} \\
   (e) \quad \dot{x} &= x^4 + x \\
   (f) \quad \dot{x} &= F(t) + G(x)
   \end{align*}
   \]

8. The following differential equations have been studied in economics. Solve them.
   \[
   \begin{align*}
   (a) \quad \dot{x} &= (At e^{a t} - b - c x) e^{x t + x} \\
   (b) \quad \dot{x} &= \frac{e^{a t} e^{a x}}{x} \quad \alpha > 0, \beta > 0, \quad a > 0, \quad a a \neq \beta \\
   \text{(Hint:} \quad \frac{x}{(\beta - a x)(x - a)} = \frac{1}{(\beta - a a) (\beta - a x + a x - a)}\text{)}
   \end{align*}
   \]
5.4 First-Order Linear Equations

A first-order linear differential equation is one that can be written in the form

$$\dot{x} + a(t)x = b(t)$$

(1)

where $a(t)$ and $b(t)$ denote continuous functions of $t$ in a certain interval, and $x = x(t)$ is the unknown function. Equation (1) is called "linear" because the left-hand side is a linear function of $x$ and $\dot{x}$.

The following are all examples of first-order linear equations:

(a) $\dot{x} + x = t$  
(b) $\dot{x} + 2tx = 4t$  
(c) $(t^2 + 1)\dot{x} + e^x = t \ln t$

The first two equations are obviously of the form (1). The last one can be put into this form if we divide each term by $t^2 + 1$ to get $\dot{x} + [(e^x)/(t^2 + 1)]x = t \ln t/(t^2 + 1)$.

The Simplest Case

Consider the following equation with $a$ and $b$ as constants, where $a \neq 0$:

$$\dot{x} + ax = b$$

(2)

Let us multiply this equation by the positive factor $e^{at}$, called an integrating factor. We then get the equivalent equation

$$\dot{x}e^{at} + axe^{at} = be^{at}$$

(3)

It may not be obvious why we came up with this idea, but it turns out to be a good one because the left-hand side of (3) happens to be the derivative of the product $xe^{at}$. Thus (3) is equivalent to

$$\frac{d}{dt}(xe^{at}) = be^{at}$$

(4)

According to the definition of the indefinite integral, equation (4) holds for all $t$ in an interval if $xe^{at} = \int be^{at}dt = (b/a)e^{at} + C$ for some constant $C$. Multiplying this equation by $e^{-at}$ gives the solution for $x$. Briefly formulated:

$$\dot{x} + ax = b \iff x = Ce^{-at} + \frac{b}{a} \quad (C \text{ is a constant})$$

(5)

If we let $C = 0$ in (5), we obtain the constant solution $x(t) = b/a$. We say that $x = b/a$ is an equilibrium state, or a stationary state, for the equation. Observe how this solution can be obtained from $\dot{x} + ax = b$ by letting $\dot{x} = 0$ and then solving the resulting equation for $x$. If the constant $a$ is positive, then the solution $x = Ce^{-at} + b/a$ converges to $b/a$ as $t \to \infty$. In this case, the equation is said to be stable, because every solution of the equation converges to an equilibrium as $t$ approaches infinity. See Section 5.7 for more on stability.

**Example 1:** Find the general solution of

$$\dot{x} + 2x = 8$$

and determine whether the equation is stable.

**Solution:** By (5), the solution is $x = Ce^{-2t} + 4$. Here the equilibrium state is $x = 4$, and the equation is stable because $a = 2 > 0$, so $x \to 4$ as $t \to \infty$.

**Example 2:** (Price adjustment mechanism) Let $D(P) = a - bP$ denote the demand and $S(P) = \alpha + \beta P$ the supply of a certain commodity when the price is $P$. Here $a$, $b$, $\alpha$, and $\beta$ are positive constants. Assume that the price $P = P(t)$ varies with time, and that $\dot{P}$ is proportional to excess demand $D(P) - S(P)$. Thus,

$$\dot{P} = \lambda[D(P) - S(P)]$$

where $\lambda$ is a positive constant. Inserting the expressions for $D(P)$ and $S(P)$ into this equation gives $\dot{P} = \lambda(a - bP - \alpha - \beta P)$. Rearranging, we then obtain

$$\dot{P} + \lambda(b + \beta)P = \lambda(a - \alpha)$$

According to (5), the solution is

$$P = Ce^{-\lambda(b+\beta)t} + \frac{a - \alpha}{b + \beta}$$

Because $\lambda(b + \beta)$ is positive, as $t$ tends to infinity, $P$ converges to the equilibrium price $P^* = (a - \alpha)/(b + \beta)$, for which $D(P^*) = S(P^*)$. Thus, the equation is stable.
6.3 Constant Coefficients

Consider the homogeneous equation

\[ x + ax + bx = 0 \]

where \( a \) and \( b \) are arbitrary constants, and \( x = x(t) \) is the unknown function. According to Theorem 6.2.1, finding the general solution of (1) requires us to discover two solutions \( u_1(t) \) and \( u_2(t) \) that are not proportional. Because the coefficients in (1) are constants, it seems a good idea to try possible solutions \( x \) with the property that \( x, \dot{x}, \dot{x}, \ddot{x} \) are all constant multiples of each other. The exponential function \( x = e^{\alpha t} \) has this property, because \( \ddot{x} = \alpha^2 x = \alpha^2 e^{\alpha t} = \alpha^2 x \) and \( \dot{x} = \alpha^2 e^{\alpha t} = \alpha^2 e^{\alpha t} \). So we try adjusting the constant \( \alpha \) in order that \( x = e^{\alpha t} \) satisfies (1). This requires us to arrange that \( r^2 \alpha^2 + ar + b = 0 \). Cancelling the positive factor \( e^{\alpha t} \) tells us that \( e^{\alpha t} \) satisfies (1) if \( r \) satisfies

\[ r^2 + ar + b = 0 \]

This is the characteristic equation of the differential equation (1). It is a quadratic equation whose roots are real if \( \frac{1}{2} a^2 - b \geq 0 \). Solving (2) by the quadratic formula in this case yields the two characteristic roots

\[ r_1 = -\frac{1}{2} a + \sqrt{\frac{1}{4} a^2 - b}, \quad r_2 = -\frac{1}{2} a - \sqrt{\frac{1}{4} a^2 - b} \]

There are three different cases which are summed up in the following theorem:

**Theorem 6.3.1**

The general solution of

\[ x + ax + bx = 0 \]

depends on the roots of the characteristic equation \( r^2 + ar + b = 0 \) as follows:

(I) If \( \frac{1}{4} a^2 - b > 0 \), when there are two distinct real roots, then

\[ x = Ae^{r_1 t} + Be^{r_2 t}, \quad \text{where} \quad r_{1,2} = -\frac{1}{2} a \pm \sqrt{\frac{1}{4} a^2 - b} \]

(II) If \( \frac{1}{4} a^2 - b = 0 \), when there is a double real root, then

\[ x = (A + Bt)e^{rt}, \quad \text{where} \quad r = -\frac{1}{2} a \]

(III) If \( \frac{1}{4} a^2 - b < 0 \), when there are no real roots, then

\[ x = e^{\alpha t}(A \cos \beta t + B \sin \beta t), \quad \alpha = -\frac{1}{2} a, \quad \beta = \sqrt{b - \frac{1}{4} a^2} \]

**Proof:**

(I) The case \( \frac{1}{4} a^2 - b > 0 \) is the simplest because there are two distinct real characteristic roots \( r_1 \) and \( r_2 \). The two functions \( e^{r_1 t} \) and \( e^{r_2 t} \) both satisfy (1), and are not proportional. So the general solution in this case is \( Ae^{r_1 t} + Be^{r_2 t} \).

(II) If \( \frac{1}{4} a^2 - b = 0 \), then \( r = -\frac{1}{2} a \) is a double root of (2), and \( u_1 = e^{rt} \) satisfies (1). We claim that \( u_2 = te^{rt} \) also satisfies (1). (See also Problem 6.) This is because \( u_2 = e^{rt} + te^{rt} \) and \( u_2 = r e^{rt} + tr e^{rt} \), which inserted into the left-hand side of (1) gives

\[ u_2 + au_2 + bu_2 = e^{rt}(a + 2r) + te^{rt}(r^2 + ar + b) \]

after simplifying. But the last expression is 0 because \( r = -\frac{1}{2} a \) and \( r^2 + ar + b = 0 \). Thus, \( e^{rt} \) and \( te^{rt} \) are indeed both solutions of equation (1). These two solutions are not proportional, so the general solution is \( Ae^{r_1 t} + Bte^{rt} \) in this case.

(III) If \( \frac{1}{4} a^2 - b < 0 \), the characteristic equation has no real roots. An example is the equation \( \ddot{x} + x = 0 \), which occurred in Problem 6.2.2; here \( a = 0 \) and \( b = 1 \), so \( \sqrt{1/4 - b} = -1 \). The general solution was \( A \sin t + B \cos t \). It should, therefore, come as no surprise that when \( \frac{1}{4} a^2 - b < 0 \), the solution of (1) involves trigonometric functions.

Define the two functions \( u_1(t) = e^{\alpha t} \cos \beta t \) and \( u_2(t) = e^{\alpha t} \sin \beta t \), where \( \alpha \) and \( \beta \) are defined in (III). We claim that both these functions satisfy (1). Since they are not proportional, the general solution of equation (1) in this case is as exhibited in (III).

Let us show that \( u_1(t) = e^{\alpha t} \cos \beta t \) satisfies (1). We find that \( u_1(t) = e^{\alpha t} \cos \beta t \beta \sin \beta t - \beta e^{\alpha t} \sin \beta t \). Furthermore, \( u_1'(t) = e^{\alpha t} \cos \beta t t - \beta e^{\alpha t} \sin \beta t \beta t \sin \beta t \beta \cos \beta t \). Hence, \( u_1 + au_1 + bu_1 = e^{\alpha t}(\beta t^2 + a \beta t + b) \cos \beta t + \beta (2a + 1) \sin \beta t \). By using the specific values of \( \alpha \) and \( \beta \), we see that \( 2a + 1 = 0 \) and \( a^2 - \beta^2 + a \beta + b = \beta = \sqrt{b - \frac{1}{4} a^2} \).