Chapter 1

Resource allocation and optimization

Economics has been defined as the science of allocating scarce resources among competing ends. Much of the microeconomic theory encountered in a first semester graduate course is concerned with the *static* allocation problem faced by firms and consumers. Techniques of constrained optimization, in particular the method of Lagrange multipliers, are employed in developing the theory of the firm and the consumer.

The optimal harvest of renewable resources or extraction of exhaustible resources is inherently a *dynamic* allocation problem; that is, the firm or resource manager is concerned with the best harvest or extraction rate through time. It turns out that the method of Lagrange multipliers can be extended to intertemporal or dynamic allocation problems in a relatively straightforward fashion. This "discrete-time" extension of the method of Lagrange multipliers serves as a useful springboard to the "continuous-time" solution of dynamic allocation problems via the maximum principle. The method of Lagrange multipliers and its various extensions reduce the original optimization problem to a system of equations to be solved. Solving this system of equations, unfortunately, can often be exceedingly difficult, especially for dynamic problems. There are also technical problems concerning sufficiency conditions for the solutions so obtained (and also pertaining to the existence of a solution to the given problem). In these notes we will normally consider problems which are simple enough that these difficulties are minimized.

Before presenting the dynamic techniques we will briefly review the method of Lagrange multipliers within the context of allocating scarce resources among competing ends at a single point in time.

1.1 Constrained optimization and the method of Lagrange multipliers

In resource economics as in other fields of economics, we often encounter constrained optimization problems. The general form of such problems is

$$\begin{aligned} &\text{maximize} \quad V(x_1, \ldots, x_n) \\ &\text{subject to} \quad (x_1, \ldots, x_n) \in A \end{aligned} \tag{1.1}$$

where $V(\cdot)$ is a given *objective* (or *value*) *function* of n decision variables $x_1, \ldots, x_n$ which are required to be in some given *constraint set* A.

In the case of a *static* optimization problem, the decision variables x_i are real numbers and the constraint set A is a subset of $\mathbb{R}^n$–Euclidean n-space. For *dynamic* optimization problems, on the other hand, some (or all) of the decision variables are functions of time t (usually separated into so-called *state* variables and *control* variables). The constraint condition then typically involves the system's *dynamics*, expressed as a system of differential or difference equations. Other constraints may also be present. As before, $V(\cdot)$ is real-valued, frequently involving integration (or summation) over time. Dynamic optimization problems will be considered in Section 1.2.

1.1.1 Static optimization: no constraints

The simplest optimization problem is

$$\text{maximize} \quad V(x_1, \ldots, x_n) \tag{1.2}$$

where the decision variables are unconstrained.[1] The reader is assumed to be familiar with the first order necessary conditions

$$\frac{\partial V(\cdot)}{\partial x_i} = 0 \qquad i = 1, \ldots, n \tag{1.3}$$

By "necessary" conditions we mean that equations (1.3) must be satisfied by the maximizing values of $x_1, \ldots, x_n$. The conditions (1.3) are not sufficient conditions for a maximum, however (they also pertain to minimal solutions and to values x_i which are neither maxima nor minima). We will not attempt to delineate sufficient conditions in these notes [see Intriligator (1971, p 26)], since such conditions often are complicated and of very limited practical use (but popular with economics professors). If the objective function $V(\cdot)$ is known to be concave, the necessary conditions (1.3) are also sufficient.[2] Note that (1.3) constitutes a system of n possibly nonlin-

[1] We shall assume throughout that $V(\cdot)$ is a smooth function; that is, all required partial derivatives exist.

[2] The function $V(X)$ is said to be concave if

$$V(\alpha \bar{X} + (1-\alpha)\tilde{X}) \geq \alpha V(\bar{X}) + (1-\alpha) V(\tilde{X})$$

for all $\bar{X} = (\bar{x}_1, \ldots, \bar{x}_n)$, $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ and $0 \leq \alpha \leq 1$.

ear equations in n unknowns $x_1, \ldots, x_n$. Thus the optimization problem has been "reduced" to the solution of n equations. Unfortunately, solving the system in practice may be almost as difficult as the original optimization problem. Numerical algorithms for the solution of such systems are available in computing centers, but do not always work (see Section 1.6.2 for an example). Frequently a direct optimization algorithm (based on a direct "search" of the feasible set) will outperform any method based on the first order necessary conditions.

Nevertheless, insight into economics is often obtained from the necessary conditions without actually solving them explicitly. For example, if $V(\cdot)$ is a net benefit function, the statement "marginal net benefit of each input x_i must equal zero" is equivalent to (1.3) and carries economic significance.

1.1.2 Static optimization: equality constraints

Consider next the constrained problem

$$\begin{aligned} &\text{maximize} \quad V(x,y,z) \\ &\text{subject to} \quad G(x,y,z) = c \end{aligned} \tag{1.4}$$

where for simplicity we have only three decision variables x, y, and z. The equation $G(\cdot) = c$, where c is a known constant, determines a constraint set in x, y, z space, which is in fact a surface, which we will denote by S_G. The problem, then, is to determine the largest value of the function $V(x,y,z)$ for points (x,y,z) on the surface S_G.

One approach to this problem is first to solve the constraint equation $G(\cdot) = c$ for one of the variables, say $z = h(x,y)$. The constrained problem in three variables may be replaced by the unconstrained two variable problem

$$\text{maximize} \quad V(x,y,h(x,y)) \tag{1.5}$$

with first order necessary conditions

$$\left.\begin{aligned} \frac{\partial V}{\partial x} + \frac{\partial V}{\partial z}\frac{\partial h}{\partial x} = 0 \\ \frac{\partial V}{\partial y} + \frac{\partial V}{\partial z}\frac{\partial h}{\partial y} = 0 \end{aligned}\right\} \tag{1.6}$$

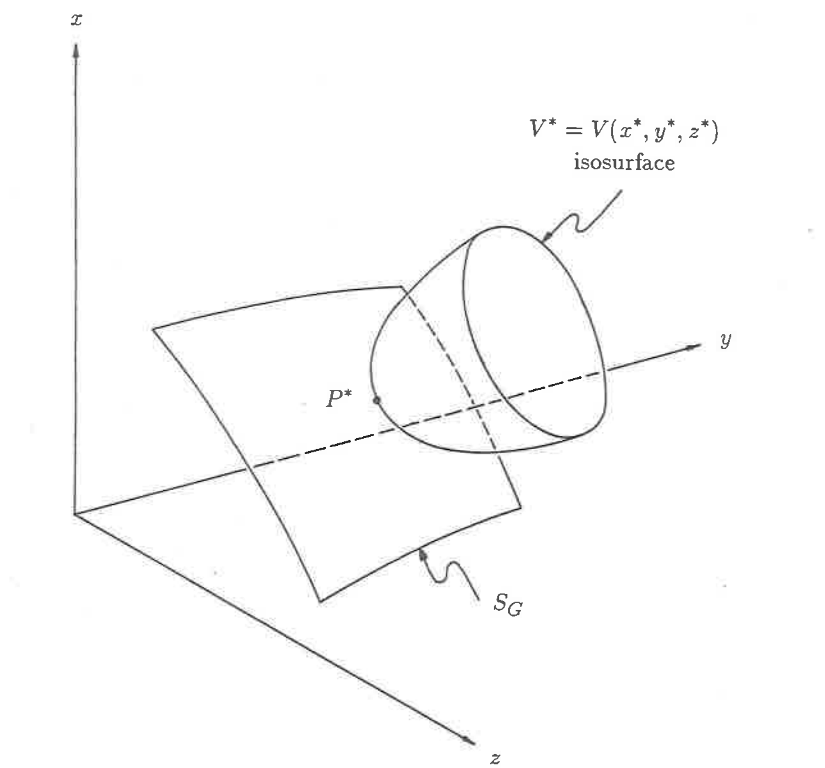


Figure 1.1 The tangency criterion: if $V(\cdot)$ is maximized on S_G at P^* then the V^* isosurface through P^* is tangent to S_G.

Differentiating the constraint equation implicitly implies

$$\left.\begin{aligned} \frac{\partial G}{\partial x} + \frac{\partial G}{\partial z}\frac{\partial h}{\partial x} = 0 \\ \frac{\partial G}{\partial y} + \frac{\partial G}{\partial z}\frac{\partial h}{\partial y} = 0 \end{aligned}\right\} \tag{1.7}$$

and Equations (1.6) can therefore be written in the form

$\frac{\partial h}{\partial z} = -\frac{V_x}{V_z}$

$$\left.\begin{aligned} V_x G_z - V_z G_x = 0 \\ V_y G_z - V_z G_y = 0 \end{aligned}\right\} \tag{1.8}$$

$G_x + G_z\left(-\frac{V_x}{V_z}\right) = 0$

where V_x is shorthand for $\partial V/\partial x$, etc.

Equations (1.8) can be obtained from an alternative geometrical derivation which gives important mathematical and economic insight. Figure 1.1 shows the solution to (1.4) as it would appear in x, y, z space. The con-

straint surface is shown as S_G. Suppose $P^* = (x^*, y^*, z^*)$ is the point on S_G for which $V(x, y, z)$ attains its maximum. Consider also the isosurface of $V(\cdot)$ passing through P^*; this is the surface

$$V(x, y, z) = V(x^*, y^*, z^*) \tag{1.9}$$

After a moment's reflection, one would conclude that $V(x^*, y^*, z^*)$ must be tangent to the constraint surface S_G. If it were not, it would either cut through S_G or not touch it at all. In the latter case the constraint is not satisfied. In the former case there would be a projection of the isosurface on S_G and there would be points on S_G lying inside and outside the projection.[3] Since $V = V^*$ on the projection we must have $V > V^*$ on one side (say inside) of the projection and $V < V^*$ on the other side (outside) of the projection. But V^* was by assumption the maximum of $V(\cdot)$ on S_G. Therefore no points of S_G can have a value $V > V^*$, which would be a contradiction. The conclusion: The maximizing isosurface must be tangent to S_G at P^*.

Recall now from calculus that the *gradient vector*

$$\vec{\nabla} V = (V_x, V_y, V_z) \tag{1.10}$$

is always perpendicular (normal) to the isosurface $V(\cdot) =$ constant at any given point. Two isosurfaces passing through a point P^* are therefore tangent at P^* if and only if their gradient vectors have the same direction. This means that $\vec{\nabla} V = \lambda \vec{\nabla} G$ for some $\lambda \neq 0$. This vector equation means, in turn, that

$$\left.\begin{aligned} V_x = \lambda G_x \\ V_y = \lambda G_y \\ V_z = \lambda G_z \end{aligned}\right\} \tag{1.11}$$

at point P^*.

Note that, by dividing pairs of equations, (1.11) reduces to the necessary conditions (1.8). Conversely (1.8) implies (1.11).[4] Equations (1.11), however, have an appealing symmetry lacking in (1.8). They also generalize in a nice way, as we shall see.

[3] The projection on an isosurface which cuts through S_G might appear as a bent circle or ellipse in Figure 1.1. Think of the "nose" of some other isosurface lying on the "other side" of S_G and the projection as a closed contour on S_G.

[4] From (1.8) we have $V_x/G_x = V_y/G_y = V_z/G_z$. Call the common value $\lambda = V_z/G_z$.

1.1.3 Lagrange multipliers

It is customary to rewrite Equations (1.11) in the form

$$\left.\begin{array}{l} V_x - \lambda G_x = 0 \\ V_y - \lambda G_y = 0 \\ V_z - \lambda G_z = 0 \end{array}\right\} \tag{1.12}$$

If we define the expression

$$L = V(x,y,z) - \lambda[G(x,y,z) - c] \tag{1.13}$$

then Equations (1.12) are also obtained as

$$L_x = L_y = L_z = 0 \tag{1.14}$$

The expression L is called the *Lagrangian* associated with the original constrained optimization problem (1.4). The number λ is referred to as a *Lagrange multiplier*. Thus, the critical observation in the development of the *method of Lagrange multipliers* was that differentiation of the Lagrangian expression would lead to the same first order necessary conditions as obtained in the simple "constraint substitution" technique used to transform an equality constrained problem into an unconstrained problem [i.e., going from problem (1.4) to problem (1.5)].

Consider now the general optimization problem with multiple equality constraints

$$\begin{array}{rl} \text{maximize} & V(x_1 \ldots, x_n) \\ \text{subject to} & G_j(x_1, \ldots, x_n) = c_j, \quad j = 1, \ldots, m \end{array} \tag{1.15}$$

The associated Lagrangian is

$$L = V(\cdot) - \sum_{j=1}^{m} \lambda_j [G_j(\cdot) - c_j] \tag{1.16}$$

Note that each of the m constraints gives rise to a Lagrange multiplier, λ_j. By the same sort of tangency argument as before it can be shown that the following equations are necessary conditions for $x_1, \ldots, x_n$ to be a solution to the above optimization problem

$$\frac{\partial L}{\partial x_i} = 0 \qquad i = 1, \ldots, n \tag{1.17}$$

Explicitly, these equations are

$$\frac{\partial V}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial G_j}{\partial x_i} = 0 \qquad i = 1, \ldots, n \tag{1.18}$$

We might also note that

$$\frac{\partial L}{\partial \lambda_j} = -G_j(\cdot) + c_j = 0 \qquad j = 1, \ldots, m \tag{1.19}$$

and that when taken together, Equations (1.18) and (1.19) constitute a system of $n+m$ equations in $n+m$ unknowns: $x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_m$. In principle this system should have at most a finite number of solutions, one of which will be the solution to our original optimization problem. In practice, as we noted earlier, solving this system of equations may be difficult indeed.

Consider the following example

$$\begin{array}{rl} \text{maximize} & 2x - 3y + z \\ \text{subject to} & x^2 + y^2 + z^2 = 9 \end{array}$$

The Lagrangian for this problem is

$$L = 2x - 3y + z - \lambda(x^2 + y^2 + z^2 - 9)$$

with the first order necessary conditions

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2 - 2\lambda x = 0 \\ \frac{\partial L}{\partial y} &= -3 - 2\lambda y = 0 \\ \frac{\partial L}{\partial z} &= 1 - 2\lambda z = 0 \\ \frac{\partial L}{\partial \lambda} &= -x^2 - y^2 - z^2 + 9 = 0 \end{aligned}$$

The first and second and first and third equations imply $y = -3x/2$ and $z = x/2$; which upon substitution into the constraint equation yields

$$x^2 + \left(\frac{-3x}{2}\right)^2 + \left(\frac{x}{2}\right)^2 = 0$$

which may be solved for $x = \pm 3\sqrt{2/7}$ leading to two solutions

$$x_1 = 3\sqrt{2/7} = 1.60 \qquad y_1 = -9/\sqrt{14} = -2.41 \qquad z_1 = 3/\sqrt{14} = 0.80$$

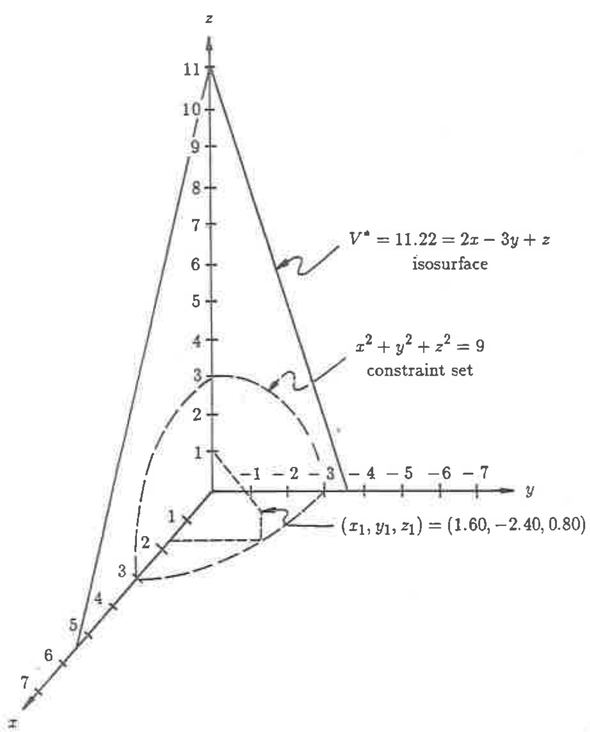


Figure 1.2 Depiction of the problem: maximize $2x - 3y + z$, subject to $x^2 + y^2 + z^2 = 9$, which has a solution at $(x_1, y_1, z_1) = (1.60, -2.41, 0.80)$.

and

$$x_2 = -3/\sqrt{2/7} = -1.60 \quad y_2 = 9/\sqrt{14} = 2.41 \quad z_2 = -3/\sqrt{14} = -0.80$$

We note, however, that

$$V(x_1, y_1, z_1) = 42/\sqrt{14} = 11.22$$
$$V(x_2, y_2, z_2) = -42/\sqrt{14} = -11.22$$

Thus, the maximizing point is (x_1, y_1, z_1), while (x_2, y_2, z_2) is a minimizing point. The problem and solution are depicted in Figure 1.2.

1.1.4 Economic interpretation

The Lagrange multipliers λ_j were not part of the original optimization problem. In the above example, for instance, we eliminated λ and then forgot about it. But Lagrange multipliers do have an important economic interpretation.

Clearly the solution to the general problem (1.15) will depend upon the values of the parameters $c_1, \ldots, c_m$ in the constraint equations $G_j(\cdot) = c_j$, $j = 1, \ldots, m$. In fact we may explicitly state this dependence as

$$x_i^* = x_i(c_1, \ldots, c_m) \tag{1.20}$$

If the optimal values for the decision variables depend on the parameters then so does the value of the objective function. Consider

$$\frac{\partial V}{\partial c_k} = \sum_{i=1}^{n} \frac{\partial V}{\partial x_i}\frac{\partial x_i}{\partial c_k} \qquad k = 1, \ldots, m \tag{1.21}$$

From Equations (1.18) we know

$$\frac{\partial V}{\partial x_i} = \sum_{j=1}^{m} \lambda_j \frac{\partial G_j}{\partial x_i} \tag{1.22}$$

and thus

$$\frac{\partial V}{\partial c_k} = \sum_{i=1}^{n}\Big(\sum_{j=1}^{m} \lambda_j \frac{\partial G_j}{\partial x_i}\Big)\frac{\partial x_i}{\partial c_k} \qquad k = 1, \ldots, m \tag{1.23}$$

Finally, differentiating the constraint equation $G_j(\cdot) = c_j$ with respect to c_k, we have

$$\sum_{i=1}^{n} \frac{\partial G_j}{\partial x_i}\frac{\partial x_i}{\partial c_k} = \delta_{j,k} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \tag{1.24}$$

Hence we find that

$$\frac{\partial V}{\partial c_k} = \sum_{j=1}^{m} \lambda_j \delta_{j,k} = \lambda_k \tag{1.25}$$

The Lagrange multiplier λ_k thus equals the incremental change in value from an incremental change in the constraint parameter c_k. In other words, λ_k represents the marginal value of relaxing the kth constraint. If c_k represents the available supply of some input or resource, then λ_k represents the "price" (or value) of the input in terms of V; hence λ_k is often called the *shadow price* of the input c_k.

Consider a resource-based economy which can allocate labor (L) to harvest timber (T) or fish (F). Assume that the economy exports both timber and fish, facing constant world prices denoted P_T and P_F, respectively. The transformation curve, relating technically efficient combinations of timber, fish, and labor, is given by

$$G(T,F;L) = T^2 + F^2/4 - L = 0$$

Suppose $P_T =$ \$500/metric ton, $P_F =$ \$1,000/metric ton and $L = 1700$ is the number of available hours of labor to be allocated between harvesting timber or fish. The static optimization problem seeks to maximize the value of harvest subject to the transformation function; that is

$$\begin{aligned} \text{maximize} \quad & V = 500T + 1{,}000F \\ \text{subject to} \quad & T^2 + F^2/4 - 1700 = 0 \end{aligned}$$

The Lagrangian expression may be written as

$$L = 500T + 1{,}000F - \lambda(T^2 + F^2/4 - 1700)$$

and has first order necessary conditions which require

$$\frac{\partial L}{\partial T} = 500 - 2\lambda T = 0$$

$$\frac{\partial L}{\partial F} = 1{,}000 - 0.5\lambda F = 0$$

and

$$\frac{\partial L}{\partial \lambda} = -T^2 - F^2/4 + 1700 = 0$$

Taking the ratio of the first two equations to eliminate λ implies $F = 8T$. Substituting this expression for F into the transformation function yields

$$\begin{aligned} T^2 + 64T^2/4 &= 1700 \\ T^2 &= 100 \end{aligned}$$

and

$$T = 10 \qquad F = 80 \qquad \lambda = 25$$

Thus, the economy should allocate the available labor so as to produce 10 metric tons of timber and 80 metric tons of fish. The marginal value (shadow price) of an additional unit of labor is \$25/hour.[5]

[5] A check of the appropriate second order conditions would reveal $T = 10$, $F = 80$, $\lambda = 25$ to be a maximum. Note: L is concave in T and F.

1.1.5 Static optimization with inequality constraints

Next let us consider the problem

$$\begin{aligned} \text{maximize} \quad & V(x,y,z) \\ \text{subject to} \quad & G(x,y,z) \le c \end{aligned} \tag{1.26}$$

The constraint set A now consists of all points lying either on the surface S_G or one particular side of S_G. There are just two possibilities: (a) the optimizing point (x,y,z) lies on one side of S_G satisfying the strict inequality $G(x,y,z) < c$ and $\partial V/\partial x = \partial V/\partial y = \partial V/\partial z = 0$, or (b) the optimizing point (x,y,z) lies on S_G, satisfying the equality $G(x,y,z) = c$ in which case the Lagrangian conditions apply and $\partial L/\partial x = \partial L/\partial y = \partial L/\partial z = 0$, where $L = V(\cdot) - \lambda[G(\cdot) - c]$.

The two cases can be combined into a single condition called the *Kuhn–Tucker condition*, which is a necessary condition, and may be written as

$$\begin{gathered} \frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = \frac{\partial L}{\partial z} = 0 \\ \lambda \begin{cases} = 0 & \text{if } G(\cdot) < c \\ \ge 0 & \text{if } G(\cdot) = c \end{cases} \end{gathered} \tag{1.27}$$

The student should check that this indeed covers cases (a) and (b) above. A frequently encountered form, equivalent to (1.27) is

$$\begin{gathered} \frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = \frac{\partial L}{\partial z} = 0 \\ \lambda[G(\cdot) - c] = 0 \\ \lambda \ge 0 \end{gathered} \tag{1.28}$$

In many applications of constrained optimization in economics the decision variables are required to be *nonnegative*; i.e., $x \ge 0$, $y \ge 0$, $z \ge 0$. If problem (1.26) is amended to include nonnegativity constraints then the Kuhn–Tucker conditions become

$$\begin{gathered} x\left(\frac{\partial L}{\partial x}\right) = y\left(\frac{\partial L}{\partial y}\right) = z\left(\frac{\partial L}{\partial z}\right) = 0 \\ x \ge 0 \qquad y \ge 0 \qquad z \ge 0 \\ \lambda[G(\cdot) - c] = 0 \qquad \lambda \ge 0 \end{gathered} \tag{1.29}$$

The Kuhn–Tucker conditions are readily generalized to the case of $x_1, \ldots, x_n$ decision variables (which may be unrestricted or nonnegative in value) plus inequality constraints $G_j(x_1, \ldots, x_n) \le c_j, j = 1, \ldots, m$. The

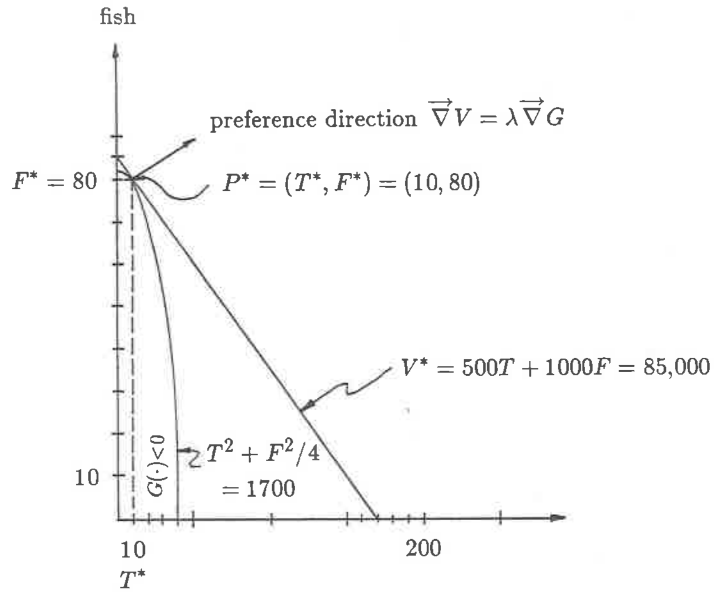


Figure 1.3 Optimal production in the timber/fish economy.

student could write out the general Kuhn–Tucker conditions as an exercise [and compare his or her version with that found in Intriligator (1971), p. 52)].

To see why $\lambda \geq 0$ let us reconsider our timber-fish economy. Suppose the transformation function were expressed as an inequality constraint; $G(T, F; L) = T^2 + F^2/4 - L \leq 0$. This constraint and the isorevenue line $V^* = 500T + 1{,}000F = 85{,}000$ are drawn in Figure 1.3. The point $P^* = (T^*, F^*) = (10, 80)$ gives the maximum revenue for $P_T = 500$, $P_F = 1{,}000$ on the constraint surface. Higher isorevenue lines would be parallel but lie to the right of V^* and thus unattainable given harvest technology and available labor.

Recall from calculus that the gradient vector, here $\vec{\nabla} V = (V_T, V_F)$, which is perpendicular to the contour $V = V^*$ also points in the direction of increasing values of V. Hence both $\vec{\nabla} V$ and $\vec{\nabla} G$ point in the same direction from P^* (namely outwards from $G(\cdot) < 0$ or inwards to $V > V^*$). Thus

$$\vec{\nabla} V = \lambda \vec{\nabla} G \tag{1.30}$$

with $\lambda > 0$.

With reference to problem (1.26) and the Kuhn–Tucker conditions as expressed in (1.27), it is possible for $\lambda = 0$ if the maximum of $V(\cdot)$ on S_G is also a *local* or *global maximum* of $V(\cdot)$ in this case $\vec{\nabla} V = 0$ so that $\lambda = 0$ in (1.30). This explains why $\lambda \geq 0$ in (1.27)–(1.29).

1.2 An extension of the method of Lagrange multipliers to dynamic allocation problems

The method of Lagrange multipliers can be employed in solving dynamic or intertemporal allocation problems and the discrete-time formulation provides a convenient introduction to control theory and the maximum principle, often presented in a continuous-time context. Let

$t = 0, 1, \ldots, T$ be the set of time periods of relevance for the dynamic allocation problem, where $t = 0$ is the present and $t = T$ is the terminal (last) period,

x_t represent a state variable, describing the system in period t,

y_t represent a control or instrument variable in period t,

$V = V(x_t, y_t, t)$ represent net economic return in period t,

$F(x_T)$ represent a final function indicating the value of alternative levels of the state variable at terminal time T, and

$x_{t+1} - x_t = f(x_t, y_t)$ be a difference equation defining the change in the state variable from period t to $(t+1)$, $t = 0, \ldots, T-1$.

The reader should note that time has been partitioned into a finite number of discrete periods, $(T+1)$ to be exact, although we can allow for an infinite horizon by letting $T \to \infty$. We will restrict ourselves to the single state, single control variable case for simplicity. The problem may be readily generalized to I state variables and J control variables. The objective function $V(\cdot)$ may have the period index t as a variable while the difference equation does not, thus $f(\cdot)$ is said to be *autonomous*.

An example of a dynamic allocation problem would be one which seeks to

$$\begin{aligned} \underset{\{y_t\}}{\text{maximize}} \quad & \sum_{t=0}^{T-1} V(x_t, y_t, t) + F(x_T) \\ \text{subject to} \quad & x_{t+1} - x_t = f(x_t, y_t) \\ & x_0 = a \quad \text{given} \end{aligned} \tag{1.31}$$

The objective in (1.31) is to maximize the sum of intermediate values plus the net value associated with the terminal state x_T. This must be done subject to the difference equation describing the change in the state variable over the horizon, assuming $x_0 = a$; that is, the initial condition is given. The problem becomes one of determining the optimal values for y_t, $t = 0, 1, \ldots, T-1$ which will, via the difference equation, imply values for x_t, $t = 1, \ldots, T$.

We can use the method of Lagrange multipliers by noting that the difference equation is a constraint equation which serves to define x_{t+1}. The Lagrangian expression may be written

$$L = \sum_{t=0}^{T-1} \{V(\cdot) + \lambda_{t+1}(x_t + f(\cdot) - x_{t+1})\} + F(\cdot) \tag{1.32}$$

where λ_{t+1} is a multiplier associated with x_{t+1}. Because there are T such constraint equations ($t = 0, \ldots, T-1$) it is appropriate to include them within the summation operation.

With no nonnegativity constraints the first order necessary conditions require:

$$\frac{\partial L}{\partial y_t} = \frac{\partial V(\cdot)}{\partial y_t} + \lambda_{t+1}\frac{\partial f(\cdot)}{\partial y_t} = 0 \qquad t = 0, \ldots, T-1 \tag{1.33}$$

$$\frac{\partial L}{\partial x_t} = \frac{\partial V(\cdot)}{\partial x_t} + \lambda_{t+1}\left(1 + \frac{\partial f(\cdot)}{\partial x_t}\right) - \lambda_t = 0 \qquad t = 1, \ldots, T-1 \tag{1.34}$$

$$\frac{\partial L}{\partial x_T} = -\lambda_T + F'(\cdot) = 0 \tag{1.35}$$

$$\frac{\partial L}{\partial \lambda_{t+1}} = x_t + f(\cdot) - x_{t+1} = 0 \qquad t = 0, \ldots, T-1 \tag{1.36}$$

Most of the partials are straightforward with the exception of (1.34) which warrants some discussion. In taking the partial of L with respect to x_t one looks at where x_t appears in the tth term of the summation. This accounts for the first two expressions on the RHS of (1.34). If, however, one were to back up to the $(t-1)$ term one would also find a $-x_t$ premultiplied by λ_t, hence the third expression $-\lambda_t$ in (1.34).

Rewriting the first order conditions will facilitate their interpretation and put them in a form more useful when making comparisons to their continuous time counterparts. They are rewritten as

$$\frac{\partial V(\cdot)}{\partial y_t} + \lambda_{t+1}\frac{\partial f(\cdot)}{\partial y_t} = 0 \qquad t = 0, 1, \ldots, T-1 \tag{1.37}$$

$$\lambda_{t+1} - \lambda_t = -\left(\frac{\partial V(\cdot)}{\partial x_t} + \lambda_{t+1}\frac{\partial f(\cdot)}{\partial x_t}\right) \qquad t = 1, \ldots, T-1 \tag{1.38}$$

$$x_{t+1} - x_t = f(\cdot) \tag{1.39}$$

$$\lambda_T = F'(\cdot) \tag{1.40}$$

$$x_0 = a \tag{1.41}$$

Equations (1.37) will typically define a marginal condition that y_t must satisfy. However, in the dynamic allocation problem there is a term not found in static problems. In many problems in resource economics $\partial V(\cdot)/\partial y_t$ will have the interpretation of being a net marginal benefit *in period* t. This is consistent with our earlier interpretation. In the dynamic context there is a second term to be accounted for in determining the optimal y_t. This term, $\lambda_{t+1}(\partial f(\cdot)/\partial y_t)$ explicitly reflects the influence of y_t on the change in the state variable. If an increase in y_t reduces the amount of *variable* x_{t+1} then this second term reflects an inter-temporal cost, often referred to as *user cost.* Less obvious, but perhaps more important, is that in the optimal solution of the problem λ^*_{t+1} can be shown to reflect the effect that an increment in x_{t+1} would have over the *remainder* of the horizon $(t+1, \ldots, T)$. Thus there is a second cost which must be considered when undertaking an incremental action today; that is, the marginal losses that might be incurred over the remaining future.

Equation (1.38) is a difference equation which must hold through time and relates the change in the Lagrange multiplier to terms involving partials of x_t. This expression can be given a nice, intuitive interpretation within the context of harvesting a renewable resource and we postpone its discussion till then. For now it is to be regarded as an equation defining how the multiplier must optimally change through time.

Equation (1.39) is simply a restatement of the difference equation for the state variable and Equations (1.40) and (1.41) are referred to as *boundary conditions* defining the terminal value of the multiplier sequence (λ_T) and the initial condition on the state variable. Because one condition is an initial condition and the other is a terminal condition, the boundary conditions are described as "split."[6]

Collectively, Equations (1.37)–(1.41) form a system of $(3T+1)$ equations in $(3T+1)$ unknowns: y_t for $t = 0, 1, \ldots, T-1$; x_t for $t = 0, 1, \ldots, T$; and λ_t for $t = 1, \ldots, T$. It may be possible to solve the system simultaneously for y_t, x_t, and λ_t although the structure for a particular problem may suggest a more efficient solution algorithm than treating it as a fully simultaneous system. If x_t, y_t, and λ_t are restricted to being nonnegative one must formulate the appropriate Kuhn–Tucker conditions and a solution might be obtained via a nonlinear programming, gradient-based algorithm.

One way of classifying dynamic problems is on the basis of whether terminal time and terminal state are given (fixed) or free to be chosen. From this perspective problem (1.31) would be classified as a "fixed-time, free-state" problem because the horizon was specified but the terminal state was not. In a free-time problem the decision-maker must determine the optimal horizon (i.e., solve for the optimal T).[7] A "restricted free-time" problem may impose a constraint on the length of horizon (e.g., $\underline{t} \le T^* \le \bar{t}$ where $\underline{t}$ and $\bar{t}$ are given). An *infinite* horizon problem, where $T \to \infty$, begs the question of whether or not the solution variables might converge to a set of values and remain unchanged thereafter. Such a solution is referred to as a *steady* or *stationary* state. If in an infinite horizon problem a steady state is attained in period τ then

$$y_t = y^*, \quad x_t = x^*, \quad \text{and} \quad \lambda_t = \lambda^* \text{ for all } t \ge \tau \tag{1.42}$$

The solution to finite (fixed) horizon problems may also lead to a stationary state. For example, it may be optimal for the manager of a mine to deplete his reserves before the end of a given planning horizon. Finally, a "terminal surface" might be specified giving the decision-maker some freedom in the selection of T and x_T, in that he must choose from permissible combinations given by $\phi(T, x_T) = 0$.

[6] It may seem to be a minor technicality that, whereas the state variable x_t is specified *initially* by Eq. (1.41), the multiplier λ_t is specified *terminally* by Eq. (1.40). However, this observation is a basic feature of dynamic optimization problems. If λ_t could also be specified initially, the system (1.39)–(1.41) could be completely solved by numerical iteration starting at $t = 0$. The fact that this cannot be done is what makes dynamic optimization difficult—and interesting!

[7] In continuous-time the optimal horizon might be determined by a differential condition $\partial L/\partial T = 0$. In discrete-time there would be no differential relationship and the decision-maker would have to explore horizons of different length, determine the optimal behavior for each horizon (T), and then compare the sum of net economic returns.

Space precludes an exhaustive discussion of the nuances of these terminal conditions. The reader is referred to Kamien and Schwartz (1981) for additional detail.

Because of its importance in many resource management situations we would like to examine more closely the infinite horizon problem and the concept of steady state. Consider the problem

$$\begin{aligned} &\underset{\{y_t\}}{\text{maximize}} \quad \sum_{t=0}^{\infty} V(x_t, y_t) \\ &\text{subject to} \quad x_{t+1} - x_t = f(x_t, y_t) \\ &\qquad\qquad x_0 = a \quad \text{given} \end{aligned} \tag{1.43}$$

In contrast to problem (1.31) the above presumes that the objective function has no explicit time dependence (t is not an argument of $V(\cdot)$) and since $T \to \infty$ there is no final function.

The Lagrangian becomes

$$L = \sum_{t=0}^{\infty} \{V(\cdot) + \lambda_{t+1}(x_t + f(\cdot) - x_{t+1})\} \tag{1.44}$$

with first order necessary conditions including:

$$\frac{\partial V(\cdot)}{\partial y_t} + \lambda_{t+1}\frac{\partial f(\cdot)}{\partial y_t} = 0 \tag{1.45}$$

$$\lambda_{t+1} - \lambda_t = -\left(\frac{\partial V(\cdot)}{\partial x_t} + \lambda_{t+1}\frac{\partial f(\cdot)}{\partial x_t}\right) \tag{1.46}$$

$$x_{t+1} - x_t = f(\cdot) \tag{1.47}$$

which must hold for $t = 0, 1, \ldots$. In steady state, with unchanging values for y_t, x_t, and λ_t, Equations (1.45)–(1.47) become a three equation system

$$\frac{\partial V(\cdot)}{\partial y} + \lambda\frac{\partial f(\cdot)}{\partial y} = 0 \tag{1.48}$$

$$\frac{\partial V(\cdot)}{\partial x} + \lambda\frac{\partial f(\cdot)}{\partial x} = 0 \tag{1.49}$$

$$f(\cdot) = 0 \tag{1.50}$$

which might be solved for the steady-state optimum y^*, x^*, and λ^*. By eliminating λ from Equations (1.48) and (1.49) and solving (1.50) for y as a function of x it is often possible to obtain a single equation in the variable x^*.

If a steady-state optimum exists for an infinite horizon problem, if it is unique, and can be found from (1.48)–(1.50), then one might ask: "If we are currently not at the steady-state optimum (i.e., $x_0 \neq x^*$), what is the best way to get there?" There are essentially two types of optimal approach paths from x_0 to x^*, assuming x^* is *reachable* from x_0. The first type is an asymptotic approach in which $x_t \to x^*$ as $t \to \infty$. The second type is called the most rapid approach path (MRAP) in which case x_t is driven to x^* as rapidly as possible, usually reaching x^* in finite time. To drive x_t to x^* as rapidly as possible will often involve a "bang-bang" control where y_t, during the MRAP assumes some maximum or minimum value.

Spence and Starrett (1975) have identified the conditions under which MRAP is optimal. The conditions for problem (1.43) are that (a) via constraint-substitution $V(x_t, y_t)$ must be expressed as an additively separable function in x_t and x_{t+1} and (b) via proper indexing, the problem may be made equivalent to optimization of $\sum_{t=1}^{\infty} w(x_t)$, where $w(\cdot)$ is quasi-concave. Interestingly enough, there are many intuitive specifications for dynamic problems which satisfy the necessary and sufficiency conditions for MRAP to be optimal. If these conditions are met, the solution of the "bang-bang" approach is a relatively trivial matter. We will give an example of such a case, shortly. Before doing so it is appropriate to introduce a more modern control theory concept: the Hamiltonian.

Look closely at conditions (1.37) to (1.39). These conditions define the dynamics between the boundary points. The *Hamiltonian* is defined as

$$\mathcal{H}(x_t, y_t, \lambda_{t+1}, t) = V(x_t, y_t, t) + \lambda_{t+1} f(x_t, y_t) \tag{1.51}$$

and it is possible to write the first order necessary conditions directly as partials of the Hamiltonian. First, note that the Lagrangian expression (1.32) may be written in terms of the Hamiltonian:

$$L = \sum_{t=0}^{T-1} \{\mathcal{H}(\cdot) + \lambda_{t+1}[x_t - x_{t+1}]\} + F(\cdot) \tag{1.52}$$

Then the first order conditions become

$$\frac{\partial L}{\partial y_t} = \frac{\partial \mathcal{H}(\cdot)}{\partial y_t} = 0 \qquad t = 0, \ldots, T-1 \tag{1.53}$$

$$\frac{\partial L}{\partial x_t} = \frac{\partial \mathcal{H}(\cdot)}{\partial x_t} + \lambda_{t+1} - \lambda_t = 0 \qquad t = 1, \ldots, T-1 \tag{1.54}$$

$$\frac{\partial L}{\partial x_T} = -\lambda_T + F'(\cdot) = 0 \tag{1.55}$$

$$\frac{\partial L}{\partial \lambda_{t+1}} = \frac{\partial \mathcal{H}(\cdot)}{\partial \lambda_{t+1}} + x_t - x_{t+1} = 0 \qquad t = 0, \ldots, T-1 \tag{1.56}$$

In their most familiar form these conditions are written as the set

$$\begin{gathered} \frac{\partial \mathcal{H}(\cdot)}{\partial y_t} = 0 \qquad \lambda_{t+1} - \lambda_t = -\frac{\partial \mathcal{H}(\cdot)}{\partial x_t} \qquad x_{t+1} - x_t = \frac{\partial \mathcal{H}(\cdot)}{\partial \lambda_{t+1}} \\ \lambda_T = F'(\cdot) \qquad x_0 = a \end{gathered} \tag{1.57}$$

The original problem, stated in (1.31), is an example of a subclass of control problems called *open-loop problems*. The solution of such a problem is a control trajectory $\{y_t^*\}$ determined as a function of time, or in our discrete-time problem, in tabular form. Knowing $\{y_t^*\}$ and x_0 one can use the difference equation $x_{t+1} = x_t + f(\cdot)$ to solve forward for the optimal trajectory x_t, denoted $\{x_t^*\}$.

Consider the following problem. As manager of a mine, you are asked to determine the optimal production schedule $\{y_t^*\}$ for $t = 0, \ldots, 9$. The mine is to be shut down and abandoned at $t = 10$. The price per unit of ore is given as $p = 1$ and the cost of extracting y_t is $c_t = y_t^2/x_t$, where x_t is *remaining reserves* at the beginning of period t.

Net revenue may be written as $\pi_t = py_t - y_t^2/x_t = [1 - y_t/x_t]y_t$ and the difference equation describing the change in remaining reserves is $x_{t+1} - x_t = -y_t$, where initial reserves are assumed given with $x_0 = 1{,}000$. Maximization of the sum of net revenues subject to reserve dynamics leads to the Hamiltonian:

$$\mathcal{H}(\cdot) = [1 - y_t/x_t]\, y_t - \lambda_{t+1} y_t$$

with the first order necessary conditions requiring:

$$\begin{aligned} \frac{\partial \mathcal{H}(\cdot)}{\partial y_t} &= 1 - 2y_t/x_t - \lambda_{t+1} = 0 && t = 0, \ldots, 9 \\ \lambda_{t+1} - \lambda_t &= -\frac{\partial \mathcal{H}(\cdot)}{\partial x_t} = -y_t^2/x_t^2 && t = 1, \ldots, 9 \\ x_{t+1} - x_t &= -y_t && t = 0, \ldots, 9 \\ x_0 &= 1{,}000, \quad \lambda_{10} = F'(\cdot) = 0 \end{aligned}$$

In this problem there is no final function and any units of x remaining in period 10 must be worthless. Note that this is a fixed-time free-state prob-

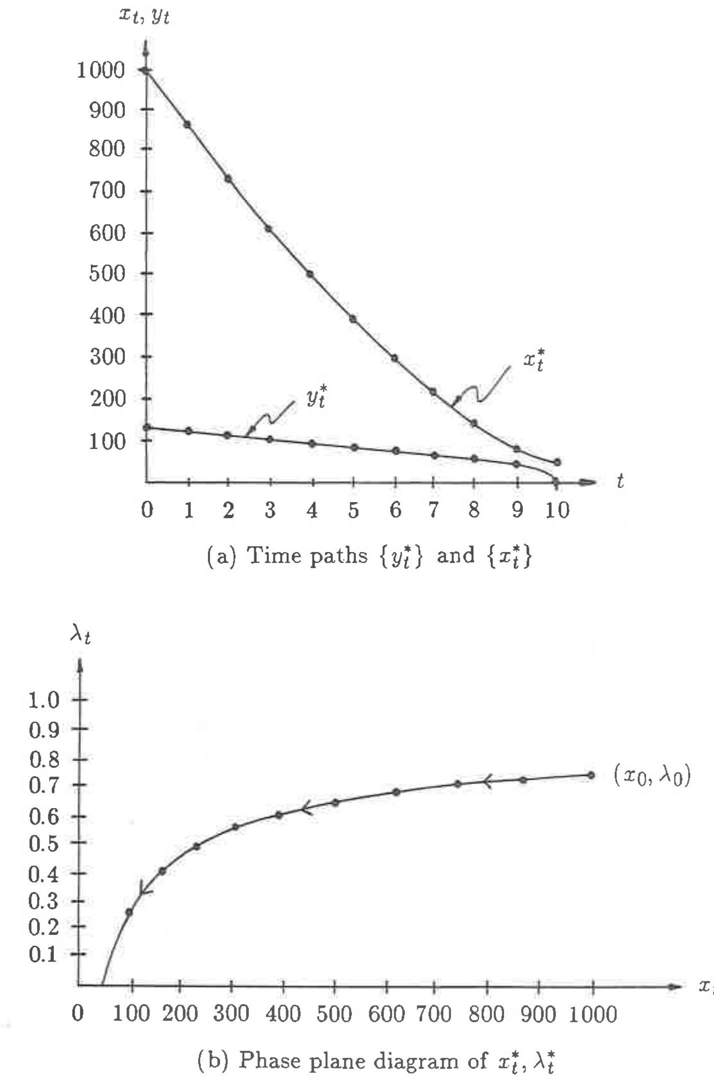
```
10 REM PROGRAM 1.1: MINE PROBLEM
20 DIM X(11),Y(11),Z(11),L(11)
30 L(10) = 0
40 FOR T = 9 TO 0 STEP -1
50    Z(T) = (1 - L(T+1)) / 2
60    L(T) = L(T+1) + Z(T)^2
70 NEXT T
80 X(0) = 1000
90 FOR T = 0 TO 9
100    Y(T) = X(T) * Z(T)
110    X(T+1) = X(T) - Y(T)
120 NEXT T
130 LPRINT " T              X(T)           Y(T)          L(T)"
140 LPRINT "-----------------------------------------------------"
150 LPRINT 0,X(0),Y(0)
160 FOR T = 1 TO 10
170    LPRINT T,X(T),Y(T),L(T)
180 NEXT T
190 END
```

T	X(T)	Y(T)	L(T)
0	1000	138.9018	
1	861.0982	129.3185	.7221965
2	731.7798	119.6851	.6996428
3	612.0947	109.993	.6728931
4	502.1016	100.2317	.6406012
5	401.8699	90.38798	.6007513
6	311.482	80.44653	.550163
7	231.0354	70.39361	.4834595
8	160.6418	60.24068	.390625
9	100.4011	50.20057	.25
10	50.20057	0	0

Program 1.1 Solution and algorithm to the mine manager's problem.

lem and that the first order conditions represent a system of 31 equations in 31 unknowns: y_t for $t = 0, 1, \ldots, 9$, x_t for $t = 0, 1, \ldots, 10$, and λ_t for $t = 1, 2, \ldots, 10$. Solution of this problem is most easily accomplished by defining $z_t = y_t / x_t$. Evaluating the $\partial \mathcal{H}(\cdot)/\partial y_t$ at $t = 9$ implies $z_9 = 0.5$ (since $\lambda_{10} = 0$). Evaluating the expression for $\lambda_{t+1} - \lambda_t$ at $t = 9$ implies $\lambda_9 = (z_9)^2 = 0.25$. Knowing λ_9 we can return to $\partial \mathcal{H}(\cdot)/\partial y_t$ to solve for z_8, then back down to the second equation for λ_8, and so forth. The last step in the recursion gives us $z_0 = 0.1389$ and $\lambda_0 = 0.7415$. Knowing that $x_0 = 1000$ we can solve for $y_0 = x_0 z_0 = 138.90$ and $x_1 = x_0 - y_0 = 861.10$. Knowing x_1 we can solve for $y_1 = x_1 z_1 = 129.32$, $x_2 = x_1 - y_1 = 731.78$, and so forth. A solution algorithm (programmed in BASIC) and the complete results are given in Program 1.1.

The optimal time paths $\{y_t^*\}$ and $\{x_t^*\}$ are plotted in Figure 1.4(a), while a plot of the point (x_t^*, λ_t^*) is shown in Figure 1.4(b). The latter



(a) Time paths $\{y_t^*\}$ and $\{x_t^*\}$

(b) Phase plane diagram of x_t^*, λ_t^*

Figure 1.4 Optimal time paths and a phase plane diagram for the mine manager's problem.

graph is referred to as a *phase plane diagram*. Arrows indicate the movement of (x_t^*, λ_t^*) over time. This simple problem can be used to illustrate other aspects of dynamic optimization problems in general and exhaustible resources in particular. We will return to this problem once more in this chapter and again in Chapter 3. We now turn to another important technique for solving dynamic optimization problems.

1.3 Dynamic programming

Consider the following problem:

$$\begin{aligned} \underset{\{y_t\}}{\text{maximize}} \quad & \sum_{t=0}^{T} V(x_t, y_t, t) \\ \text{subject to} \quad & x_{t+1} = f(x_t, y_t, t), \quad t = 0, 1, \ldots, T-1 \qquad (1.58) \\ & y_t \in Y, \quad t = 0, 1, \ldots, T \\ & x_0 = a \text{ given} \end{aligned}$$

In comparison to Problem (1.31) the final function has been incorporated as $V(x_T, y_T, T)$, all x_t terms have been collected on the right hand side of the difference equation, and y_t must come from the set Y, for $t = 0, \ldots, T$ (as opposed to $t = 0, \ldots, T-1$). However, these changes are mainly for notational convenience only.

Define $J_n(x)$ as the maximum total value when only n periods remain, and the state variable at the outset of these n periods is x. Thus

$$J_n(x) = \max \sum_{t=T-(n-1)}^{T} V(x_t, y_t, t), \quad \text{given } x_{T-(n-1)} = x \qquad (1.59)$$

subject to the same constraints as in the original problem for $t \geq T-(n-1)$.

For $n = 1$ we have simply

$$J_1(x) = \max_{y_T \in Y} V(x_T, y_T, T), \quad x_T = x \qquad (1.60)$$

that is, a single constrained static optimization problem. Suppose this problem to have been solved for every value of x in a range of interest. Next consider $n = 2$ and

$$\begin{aligned} J_2(x) = \max_{Y_{T-1} \in Y} & [V(x_{T-1}, y_{T-1}, T-1) \\ & + J_1(f(x_{T-1}, y_{T-1}, T-1))] \qquad (1.61) \\ x_{T-1} = x & \end{aligned}$$

This equation is easily explained: Let some decision on y be adopted in period $T-1$. The first term on the right of (1.61) is the immediate payoff. After this decision is made, only one period remains and the state of the system will be $x_T = f(x_{T-1}, y_{T-1}, T-1)$ as per our difference equation in

Problem (1.58). If the entire policy is optimal *this final decision must also be taken optimally.* But the optimal terminal period value is

$$J_1(x_T) = J_1(f(x_{T-1}, y_{T-1}, T-1)) \qquad (1.62)$$

Finally, y_{T-1} itself must be chosen optimally so that (1.61) is valid.

Now observe that if $J_1(x)$ is assumed known for all x, (1.61) is again a static optimization problem. Having solved this problem we can continue:

$$\begin{aligned} J_3(x) = \max_{Y_{T-2} \in Y} & [V(x_{T-2}, y_{T-2}, T-2) \\ & + J_2(f(x_{T-2}, y_{T-2}, T-2))] \qquad (1.63) \\ x_{T-2} = x & \end{aligned}$$

and so on, until $n = T+1$ and the original problem has been solved. This is the famous method of *dynamic programming* (Bellman 1957), sometimes called "backwards induction"—for obvious reasons. The general expression for n "periods to go" is

$$\begin{aligned} J_n(x) = \max_{Y_{T-(n-1)}} & [V(x_{T-(n-1)}, y_{T-(n-1)}, T-(n-1)) \\ & + J_{n-1}(f(x_{T-(n-1)}, y_{T-(n-1)}, T-(n-1)))] \qquad (1.64) \\ x_{T-(n-1)} = x & \end{aligned}$$

and is called "Bellman's equation." The argument giving rise to it is the "principle of optimality" which states:

> An optimal policy has the property that whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision. (Bellman, 1957)

What practical difficulties might arise in using dynamic programming? Each of the static optimization problems must be solved for all the relevant values of the initial variable x. In cases where x is multidimensional this problem may be highly nontrivial—Bellman called this the "curse of dimensionality." Nevertheless, dynamic programming has found wide application.

As an example, suppose we sought to

$$\underset{\{y_t\}}{\text{maximize}} \quad V = \sum_{t=0}^{4} [10x_t - 0.1y_t^2]$$

$$\text{subject to} \quad x_{t+1} = x_t + y_t$$
$$x_0 = 0, \quad y_t \geq 0$$

We begin by noting that the value function with one period to go is

$$J_1 = \max_{y_4 \geq 0} [10x_4 - 0.1y_4^2]$$

Since a positive level for y_4 will only reduce J_1 we quickly determine $y_4^* = 0$ and $J_1 = 10x_4$. Proceeding backwards

$$\begin{aligned} J_2 &= \max_{y_3 \geq 0} [10x_3 - 0.1y_3^2 + J_1] \\ &= \max_{y_3 \geq 0} [10x_3 - 0.1y_3^2 + 10x_4] \\ &= \max_{y_3 \geq 0} [10x_3 - 0.1y_3^2 + 10(x_3 + y_3)] \end{aligned}$$

Taking the partial derivative of the last bracketed expression with respect to y_3 requires $-0.2y_3 + 10 = 0$, or $y_3 = 50$. Substituting this value back into the bracketed expression produces $J_2 = 20x_3 + 250$.

Continuing backwards

$$\begin{aligned} J_3 &= \max_{y_2 \geq 0} [10x_2 - 0.1y_2^2 + J_2] \\ &= \max_{y_2 \geq 0} [10x_2 - 0.1y_2^2 + 20x_3 + 250] \\ &= \max_{y_2 \geq 0} [10x_2 - 0.1y_2^2 + 20(x_2 + y_2) + 250] \end{aligned}$$

Taking the partial of $[\cdot]$ with respect to y_2 requires $-0.2y_2 + 20 = 0$ or $y_2 = 100$ and thus $J_3 = 30x_2 + 1,250$.

Proceeding in a similar fashion back to $J_5 = -0.1y_0^2 + 40y_0 + 3500$ the following solution obtains:

t	y_t	x_t
0	200	0
1	150	200
2	100	350
3	50	450
4	0	500

$$V = J_5 = 7,500$$

1.4 Continuous-time problems and the maximum principle

When time is taken as continuous the optimization interval becomes $0 \leq t \leq T$ and the difference equation describing the change in the state variable is replaced by the differential equation $dx(t)/dt = \dot{x} = f(\cdot)$. The continuous-time analogue to problem (1.31) is

$$\begin{aligned} &\text{maximize} \quad \int_0^T V(x(t), y(t), t)\, dt + F(x(T)) \\ &\text{subject to} \quad \dot{x} = f(x(t), y(t)) \qquad (1.65) \\ &\qquad x(0) = a \quad \text{given} \end{aligned}$$

Note that the integration operator has replaced the discrete-time summation operator and that, by convention, the continuous-time variables parenthesize t as opposed to subscripting. In the continuous-time problem it is necessary to assume $x(t)$ is continuous and $y(t)$ piecewise continuous. In the following development, note the close analogy with the discrete-time problem discussed above.

Proceeding as before we may form a Lagrangian expression

$$L = \int_0^T [V(\cdot) + \lambda(t)(f(\cdot) - \dot{x})]\, dt + F(\cdot) \qquad (1.66)$$

The term $-\lambda(t)\dot{x}$ may be integrated by parts to yield $\int_0^T \dot{\lambda} x(t)\, dt - [\lambda(T)x(T) - \lambda(0)x(0)]$ which upon substitution into (1.66) gives

$$L = \int_0^T [V(\cdot) + \lambda(t)f(\cdot) + \dot{\lambda}x(t)]\, dt + F(\cdot) - [\lambda(T)x(T) - \lambda(0)x(0)] \qquad (1.67)$$

Defining the continuous-time Hamiltonian as

$$\mathcal{H}(x(t), y(t), \lambda(t), t) = V(x(t), y(t), t) + \lambda(t)f(x(t), y(t)) \qquad (1.68)$$

we may rewrite the Langrangian expression yet again as

$$L = \int_0^T [\mathcal{H}(\cdot) + \dot{\lambda}x(t)]\, dt + F(\cdot) - [\lambda(T)x(T) - \lambda(0)x(0)] \qquad (1.69)$$

The first order necessary condition may be derived in the following heuristic manner. Consider a change in the control trajectory from $y(t)$ to $y(t) + \Delta y(t)$ which causes a change in the state trajectory from $x(t)$ to

$x(t) + \Delta x(t)$. The change in the Lagrangian is:

$$\Delta L = \int_0^T \left[\frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} \Delta y(t) + \frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} \Delta x(t) + \dot{\lambda} \Delta x(t) \right] dt + \left[F'(\cdot) - \lambda(T) \right] \Delta x(T) \tag{1.70}$$

For a maximum it is necessary that the change in the Lagrangian vanish for any $\{\Delta y(t)\}$, thus

$$\begin{gathered} \frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} = 0 \\ \dot{\lambda} = -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} \\ \lambda(T) = F'(\cdot) \end{gathered} \tag{1.71}$$

✓

From the definition of $\mathcal{H}(\cdot)$, and taking into account the initial condition, we may write the necessary conditions in their entirety as

$$\begin{gathered} \frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} = 0 \qquad \dot{\lambda} = -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} \qquad \dot{x} = \frac{\partial \mathcal{H}(\cdot)}{\partial \lambda(t)} \\ \lambda(T) = F'(\cdot) \qquad x(0) = a \end{gathered} \tag{1.72}$$

The reader should compare conditions (1.72) with their discrete-time analogues given in (1.57). In both discrete- and continuous-time we might summarize as follows:

(a) $x_t, x(t)$ is the state variable.
(b) $y_t, y(t)$ is the control variable.
(c) $\lambda_t, \lambda(t)$ is the adjoint or costate variable.
(d) $x_{t+1} - x_t = f(\cdot)$, $\dot{x} = f(\cdot)$ is the state equation or equation of motion.
(e) $\partial \mathcal{H}(\cdot)/\partial y_t = 0$, $\partial \mathcal{H}(\cdot)/\partial y(t) = 0$ is the maximum condition.
(f) $\lambda_{t+1} - \lambda_t = -\partial \mathcal{H}(\cdot)/\partial x_t$, $\dot{\lambda} = -\partial \mathcal{H}(\cdot)/\partial x(t)$ is the adjoint equation.

As before, alternative terminal conditions may be considered. For example, suppose $x(T) = b$ is specified. Then the last term in (1.70) disappears (because $\Delta x(T) = 0$) so that the last equation in (1.71) is no longer valid. (The total number of terminal conditions on $x(T)$ and $\lambda(T)$ always remains the same.) If terminal time is free we must have $\partial L/\partial T = 0$. From (1.69) we obtain

$$\frac{\partial L}{\partial T} = \mathcal{H}(T) + \dot{\lambda}(T)x(T) + F'(\cdot)\dot{x}(T) - \dot{\lambda}(T)x(T) - \lambda(T)\dot{x}(T) = \mathcal{H}(T) = 0$$

from the last condition in (1.71). Thus, the free terminal-time condition

is simply

$$\mathcal{H}(T) = \mathcal{H}(x(T), y(T), \lambda(T), T) = 0 \tag{1.73}$$

which along with $\lambda(T) = F'(\cdot)$, when it applies, are referred to as the *transversality conditions.*

Finally, the conditions

$$\begin{gathered} \frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} = 0 \\ \dot{\lambda} = -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} \qquad \lambda(T) = F'(\cdot) \\ \mathcal{H}(x(T), y(T), \lambda(T), T) = 0 \end{gathered} \tag{1.74}$$

are collectively referred to as the *maximum principle.* The above form of the maximum principle, which was part of the calculus of variations, does not pertain to the case where control constraints, such as $y(t) \in Y$, exist. In this case, it has been proved (Pontryagin et al. 1962) that the maximum condition, $\partial \mathcal{H}(\cdot)/\partial y(t) = 0$, takes the more general (and more incisive) form:

$$\begin{gathered} y(t) \quad \text{maximizes} \quad \mathcal{H}(x(t), y, \lambda(t), t) \\ \text{over} \quad y(t) \in Y \quad \text{for} \quad 0 \le t \le T \end{gathered} \tag{1.75}$$

If $y(t)$ is an interior point of the control set Y then (1.75) reduces to $\partial \mathcal{H}(\cdot)/\partial y(t) = 0$. This apparently minor generalization has had a profound effect for the theory and applications of dynamic optimization. Nowadays, the phrase "maximum principle" is usually taken to imply the more general maximum condition (1.75).[8]

The economic interpretation of $\lambda(t)$ is facilitated by defining the value function

$$J(x,t) = \max_{\{y(t)\}} \int_t^T V(x(t), y(t), t)\, dt \tag{1.76}$$

for $\dot{x} = f(\cdot)$, $y(t) \in Y$, and $x(t) = x$ (given). In analogy to the discrete-time case it can be shown that, for the optimal solution

$$\lambda(t) = \partial J/\partial x \tag{1.77}$$

that is, the shadow price $\lambda(t)$ equals the marginal value of the state variable at time t.

[8] This condition has been proven to hold for discrete-time problems; see Cannon, Cullum, and Polak (1970).

The Hamiltonian

$$\mathcal{H}(t) = \mathcal{H}(x(t), \lambda(t), t) = V(\cdot) + \lambda(t) f(\cdot) \tag{1.78}$$

can now be interpreted as the total rate of increase in the value of assets: The first term, $V(\cdot)$, is the flow of net returns at instant t while the second term, $\lambda(t)f(\cdot)$, is the increase in the value of the stock, x. With this interpretation, the maximum condition (1.75) makes eminent sense—the optimal control maximizes the total rate of increase of value of assets.

As an example, suppose a renewable resource may be costlessly harvested and sold at the price $p(t)$. Let $h(t)$ denote the rate of harvest and assume that capital or other fixed factors limit harvest so that $0 \le h(t) \le \bar{h}$. If $x(t)$ is the stock or biomass of the renewable resource, assume that $\dot{x} = F(x) - h(t)$ is the change in the stock where $F(x)$ is a function defining the net natural growth rate and $x(0)$ is given.[9] The problem for a resource manager or owner may be to

$$\begin{aligned} \text{maximize} \quad & \int_0^T p(t)h(t)\,dt \\ \text{subject to} \quad & \dot{x} = F(x) - h(t) \\ 0 \le h(t) \le \bar{h}, \quad & x(0) \quad \text{given} \end{aligned}$$

The Hamiltonian for this problem is

$$\begin{aligned} \mathcal{H}(x(t), y(t), \lambda(t), t) &= p(t)h(t) + \lambda(t)[F(x) - h(t)] \\ &= h(t)[p(t) - \lambda(t)] + \lambda(t)F(x) \end{aligned}$$

The maximum condition says that $h(t)$ must maximize this expression, and since $\mathcal{H}(\cdot)$ is linear in harvest this implies

$$h(t) = \begin{cases} 0 & \text{if } p(t) < \lambda(t) \\ \bar{h} & \text{if } p(t) > \lambda(t) \end{cases}$$

The most important case, however, is when $p(t) = \lambda(t)$. In this case we know $\dot{p} = \dot{\lambda}$. But from the adjoint equation we also know

$$\dot{\lambda} = -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} = -\lambda(t)F'(\cdot)$$

or

$$F'(\cdot) = -\dot{p}/p(t)$$

[9] The function $F(x)$ will be discussed in greater detail in Chapter 2.

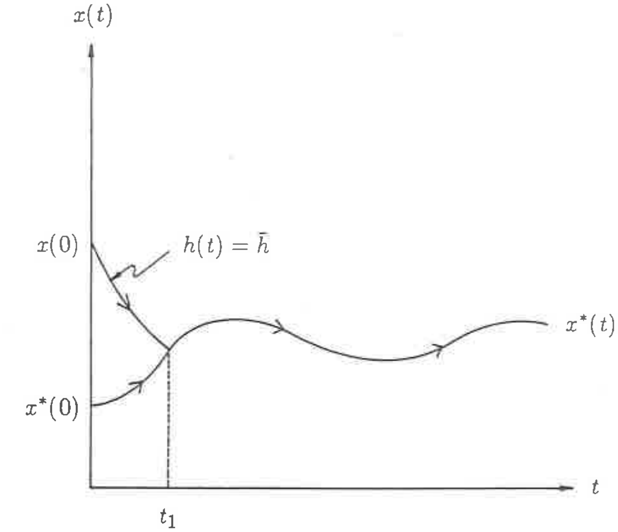


Figure 1.5 Optimal time path for $x(t)$ obtained from the solution of $F'(\cdot) = -\dot{p}/p(t)$.

This equation implies that the biomass of the resource should be maintained so that the change in the net natural growth rate equals the negative of the percentage rate of change in price.[10] With $p(t)$ known and continuous this equation will imply an optimal time path $x^*(t)$ (assuming $F(\cdot) = -\dot{p}/p(t)$ can be solved for $x^*(t)$). Such a path is drawn in Figure 1.5.

From the maximum principle we have learned that either the optimal solution must coincide with $x^*(t)$, or else $h(t) = 0$ or $\bar{h}$. Clearly the complete solution must be given by

$$h(t) = \begin{cases} 0 & \text{if } x(t) < x^*(t) \\ h^*(t) = F(x^*(t)) - \dot{x}^*(t) & \text{if } x(t) = x^*(t) \\ \bar{h} & \text{if } x(t) > x^*(t) \end{cases}$$

This is an example of the Most Rapid Approach Path (MRAP) or a "bang-bang" control, where, in this case, harvest takes place at its lower or upper bound to return stock to its optimal level as rapidly as possible.[11] In Figure 1.5 if the initial stock $x(0) > x^*(0)$, then $h(t) = \bar{h}$ until $x(t) = x^*(t)$ at $t = t_1$. The solution $x^*(t)$ is called a *singular* solution. Such solutions arise in control problems in which the Hamiltonian expression is *linear* in

[10] If revenues were discounted at rate δ the rule would be $F'(\cdot) = \delta - \dot{p}/p(t)$. Discounting will be discussed more fully in the next section.
[11] There could also be a "bang-bang" adjustment to the terminal value $x(T)$ if $T < \infty$.

the control variable. The coefficient of $h(t)$ in the Hamiltonian, in this case,

$$\sigma(t) = p(t) - \lambda(t)$$

is called a *switching function* because it determines whether $h(t)$ should switch from $h = \bar{h}$ to $h = 0$ or $h = h^*(t)$; the latter case occurring when $\sigma(t) \equiv 0$.

In economics and operations research, singular solutions are often called *myopic* solutions. Why? In determining the optimal resource stock at a particular point in time we only needed to know the current price $p(t)$ and the rate of change in price $\dot{p}(t)$. Future price changes (i.e., the complete time path $p(t)$) need not be known to formulate the optimal policy! While this is an encouraging result for planners it must be realized that it will hold only under special assumptions. Namely, that there must be zero adjustment costs and the price must not fluctuate so rapidly that the myopic path $x^*(t)$ becomes infeasible (see Clark 1976, Chapter 2).

As a second example, consider the following nonlinear, free-time, free-state problem[12]

$$\text{maximize} \quad V = \int_0^T \frac{a}{b}(1 - e^{-by(t)})\gamma(t)\,dt$$
$$\text{subject to} \quad \dot{x}(t) = -y(t)$$
$$x(0) \quad \text{given} \qquad x(t) \geq 0$$

where $a > 0$, $b > 0$, $\gamma(t) > 0$ and to set the stage for discounting suppose $\dot{\gamma}/\gamma(t) = -\delta < 0$, where $\delta > 0$. The Hamiltonian for this problem may be written as

$$\mathcal{H}(\cdot) = \frac{a}{b}\left(1 - e^{-by(t)}\right)\gamma(t) - \lambda(t)y(t)$$

and the maximum principle requires

$$\frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} = ae^{-by(t)}\gamma(t) - \lambda(t) = 0$$

$$\dot{\lambda} = -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} = 0$$

The adjoint equation implies $\lambda(t) = \lambda$, a constant. Taking the time derivative of the maximum condition, we obtain

$$-abe^{-by(t)}\dot{y}\gamma(t) + ae^{-by(t)}\dot{\gamma} = \dot{\lambda} = 0$$

[12] This can be thought of as a simplistic exhaustible resource model.

which simplifies to

$$\dot{y} = \frac{\dot{\gamma}}{b\gamma(t)} = -\frac{\delta}{b} \quad (\text{recall } \dot{\gamma}/\gamma = -\delta)$$

This last expression may be integrated directly to imply

$$y(t) = -\frac{\delta}{b}t + c_0$$

where c_0 is a constant of integration. The *transversality condition* takes the form:

$$\mathcal{H}(T) = \frac{a}{b}(1 - e^{-by(T)})\gamma(t) - \lambda y(T) = 0$$

which will only hold for $y(T) = 0$. Evaluating $y(t)$ at $t = T$ gives

$$y(T) = -\frac{\delta T}{b} + c_0 = 0, \quad \text{or} \quad c_0 = \frac{\delta T}{b}$$

Thus,

$$y(t) = \frac{\delta}{b}(T - t)$$

If $x(T) > 0$ it would be possible to increase $y(t)$ which would increase V so in this case it is also optimal for $x(T) = 0$, which in turn implies

$$x(0) = \int_0^T y(t)\,dt = \int_0^T \frac{\delta}{b}(T - t)\,dt$$
$$= \frac{\delta}{b}\left(Tt - \frac{t^2}{2}\right)\Big|_0^T = \frac{\delta T^2}{2b}$$

Solving for T yields

$$T = \sqrt{2bx(0)/\delta}$$

Given specific values for b, $x(0)$, and δ one can calculate T^*, $y^*(t)$, and $x^*(t)$.

1.5 Discounting

Discounting is a technique for calculating the *present value* of a future stream of net income. In discrete-time models we could calculate the present value of future net incomes N_t, $t = 0, 1, 2, \ldots, T$ according to

$$N = \sum_{t=0}^{T} N_t/(1+\delta)^t = \sum_{t=0}^{T} \rho^t N_t \qquad (1.79)$$

where $\rho = 1/(1+\delta)$ is the discount *factor* and δ is the periodic *discount rate.* In searching for the "best" dynamic allocation one frequently encounters objective functions that represent a present value.

In continuous time the present value of net incomes $N(t)$, $0 \le t \le T$ may be calculated according to

$$N = \int_0^T N(t)e^{-rt}\,dt \tag{1.80}$$

In Equation (1.80) the term e^{-rt} is the continuous (or instantaneous) discount factor and r is the instantaneous rate of discount. Assuming that the time units are the same in (1.79) and (1.80), we see that

$$e^{-r} = \frac{1}{1+\delta}, \qquad \text{i.e.,} \qquad r = \ln(1+\delta)$$

For example, a 10% discount rate compounded annually is equivalent to a continuous rate of 9.53%. Other compounding periods may be treated by a similar calculation.

For varying discount rates we have the following formulas for the discount factor:

$$\text{discrete time:} \qquad \rho(t) = \frac{1}{1+\delta_1}\cdot\frac{1}{1+\delta_2}\cdot\dots\cdot\frac{1}{1+\delta_t}$$

$$\text{continuous time:} \qquad \rho(t) = \exp\Big(-\int_0^t r(s)\,ds\Big)$$

The reader should check that these formulas reduce to the previous expressions in the case of constant discount rates.

We now consider the form of the discrete- and continuous-time problems with present value objectives.

Consider the discrete-time problem

$$\begin{aligned} &\text{maximize} \sum_{t=0}^{T-1} \rho^t V(x_t, y_t) + \rho^T F(x_T) \\ &\text{subject to} \quad x_{t+1} - x_t = f(x_t, y_t) \\ &\qquad x_0 = a \quad \text{given} \end{aligned} \tag{1.81}$$

There are two ways one might formulate the Lagrangian expression for this problem. We take the following approach because it offers more straightforward interpretations. Define

$$L = \sum_{t=0}^{T-1} \rho^t \{V(\cdot) + \rho\lambda_{t+1}[x_t + f(\cdot) - x_{t+1}]\} + \rho^T F(\cdot) \tag{1.82}$$

In (1.82) we premultiply the difference equation by λ_{t+1} as before but then premultiply the product by the discrete-time discount factor ρ. The multiplier λ_{t+1} can be interpreted as the value of an additional unit of x_{t+1} from the *perspective of period* $t+1$. The term $\lambda_{t+1}[\cdot]$ is thus a value in period $t+1$. The objective function $V_t = V(x_t, y_t)$, however, represents a value in period t. To make these two values comparable we discount $\lambda_{t+1}[\cdot]$ one period, thus the expression in $\{\cdot\}$ is a value from the perspective of period t, and it is discounted by ρ^t, then added to the discounted values from other periods to calculate the present value of the Lagrangian.

The discrete-time *current value* Hamiltonian is

$$\tilde{\mathcal{H}}(x_t, y_t, \lambda_{t+1}) = V(\cdot) + \rho\lambda_{t+1} f(\cdot) \tag{1.83}$$

and the first order conditions will require

$$\begin{gathered} \frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial y_t} = 0 \\ \rho\lambda_{t+1} - \lambda_t = -\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial x_t} \qquad x_{t+1} - x_t = \frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial(\rho\lambda_{t+1})} \\ \lambda_T = F'(\cdot) \qquad x_0 = a \end{gathered} \tag{1.84}$$

The above conditions can be verified by rewriting the Lagrangian in terms of the discrete-time current value Hamiltonian. The term "current value" stems from the fact that the Hamiltonian as we have defined it represents a value from the perspective of period t. In comparing conditions (1.57) without discounting to (1.84) with discounting note the discount factor which premultiplies λ_{t+1} in (1.84).

Suppose, if instead of the finite horizon problem in (1.81), we had the infinite horizon problem

$$\begin{aligned} &\text{maximize} \sum_{t=0}^{\infty} \rho^t V(x_t, y_t) \\ &\text{subject to} \quad x_{t+1} - x_t = f(x_t, y_t) \\ &\qquad x_0 = a \quad \text{given} \end{aligned} \tag{1.85}$$

The current value Hamiltonian is again given by (1.83) and the first order conditions will imply

$$\frac{\partial V(\cdot)}{\partial y_t} + \rho\lambda_{t+1}\frac{\partial f(\cdot)}{\partial y_t} = 0 \tag{1.86}$$

$$\rho\lambda_{t+1} - \lambda_t = -\frac{\partial V(\cdot)}{\partial x_t} - \rho\lambda_{t+1}\frac{\partial f(\cdot)}{\partial x_t} \tag{1.87}$$

$$x_{t+1} - x_t = f(\cdot) \tag{1.88}$$

Suppose a steady state exists and is reachable from $x_0 = a$. Then evaluating (1.86) at steady state implies $\lambda = -(1+\delta)[V_y/f_y]$. Substituting this expression into (1.87) (also evaluated at steady state) and isolating δ on the RHS yields

$$f_x - \frac{V_x}{V_y} f_y = \delta \tag{1.89}$$

Equation (1.89) is a fundamental result to models of renewable resources and in the next chapter it will be given a capital-theoretic interpretation. Together with Equation (1.88), which evaluated at a steady state implies $f(\cdot) = 0$, we obtain a two equation system that may be solved for the steady-state optimum (x^*, y^*).

Let us next reconsider the mine manager's problem posed in Section 1.2, but now introduce discounting. The problem becomes:

$$\begin{aligned} \text{maximize}\, V &= \sum_{t=0}^{9} \rho^t [1 - y_t/x_t] y_t \\ \text{subject to} \quad x_{t+1} - x_t &= -y_t \\ x_0 &= 1,000 \end{aligned}$$

The current value Hamiltonian is

$$\tilde{\mathcal{H}}(\cdot) = [1 - y_t/x_t]y_t - \rho\lambda_{t+1}y_t$$

and the first order conditions require

$$\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial y_t} = 1 - 2y_t/x_t - \rho\lambda_{t+1} = 0 \qquad t = 0, 1, \ldots, 9$$

Table 1.1 *Solution to the mine manager's problem with discounting* ($\delta = 0.10$).

t	x_t	y_t	λ_t
0	1,000	238.23	–
1	761.77	183.68	0.5759
2	578.09	141.82	0.5695
3	436.27	109.71	0.5603
4	326.56	85.09	0.5467
5	241.47	66.22	0.5267
6	175.25	51.74	0.4967
7	123.51	40.61	0.4505
8	82.90	32.03	0.3765
9	50.87	25.43	0.2500
10	25.44	0	0

$$\rho\lambda_{t+1} - \lambda_t = \overset{(-)}{\wedge}\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial x_t} = -y_2^2/x_t^2 \qquad t = 1, 2, \ldots, 9$$

$$x_{t+1} - x_t = -y_t \qquad t = 0, 1, \ldots, 9$$

$$x_0 = 1,000, \qquad \lambda_{10} = F'(\cdot) = 0$$

The solution algorithm is the same, only now the discount factor will influence the values of $z_8, \ldots, z_0$; $\lambda_8, \ldots, \lambda_0$, and will thus affect the optimal values of y_t, $t = 0, 1, \ldots, 9$, and x_t, $t = 1, 2, \ldots, 10$, when solving forward for these values. If $\delta = 0.10$ the values for x_t and y_t are shown in Table 1.1. In each case, price is assumed constant, $p = 1$, and initial reserves are $x_0 = 1,000$. With discounting, the manager finds it optimal to produce more initially, driving remaining reserves down more quickly and producing less terminally (e.g., compare y_8 in Program 1.1 and Table 1.1). In this problem, discounting leads to greater production over the entire horizon (note $x_{10} = 25.4$ when $\delta = 0.1$). This tilting of production profiles toward the present (and the more rapid depletion of nonrenewable resources) is an all-but-universal feature of discounting. The higher the discount rate, the greater the tilt.

Discounting in the continuous-time model is also facilitated by introducing the current value Hamiltonian. Consider the following problem.

$$\begin{aligned} \text{maximize} \int_0^T V(x(t),y(t))e^{-\delta t}\,dt + F(x(T))e^{-\delta T} \\ \text{subject to} \quad \dot{x} = f(x(t),y(t)) \\ x(0) = a \quad \text{given} \end{aligned} \tag{1.90}$$

The Hamiltonian for this problem is

$$\mathcal{H}(\cdot) = V(\cdot)e^{-\delta t} + \lambda(t)f(\cdot) \tag{1.91}$$

The current value Hamiltonian is defined as

$$\tilde{\mathcal{H}} = \mathcal{H}(\cdot)e^{\delta t} = V(\cdot) + \mu(t)f(\cdot) \tag{1.92}$$

where $\mu(t) = e^{\delta t}\lambda(t)$. The necessary conditions given in (1.72) are in terms of the Hamiltonian and require, in part, that

$$\frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} = \frac{\partial V(\cdot)}{\partial y(t)}e^{-\delta t} + \lambda(t)\frac{\partial f(\cdot)}{\partial y(t)} = 0 \tag{1.93}$$

$$\begin{aligned} \dot{\lambda} &= -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} \\ &= -\frac{\partial V(\cdot)}{\partial x(t)}e^{-\delta t} - \lambda(t)\frac{\partial f(\cdot)}{\partial x(t)} \end{aligned} \tag{1.94}$$

From the definition for $\mu(t)$ we note $\lambda(t) = \mu(t)e^{-\delta t}$ and $\dot{\lambda} = -\delta\mu(t)e^{-\delta t} + \dot{\mu}e^{-\delta t}$. Equations (1.93) and (1.94) may be rewritten in terms of $\mu(t)$ and $\dot{\mu}$ so that

$$\frac{\partial V(\cdot)}{\partial y(t)} + \mu(t)\frac{\partial f(\cdot)}{\partial y(t)} = 0 \tag{1.95}$$

$$\dot{\mu} = -\frac{\partial V(\cdot)}{\partial x(t)} + \mu(t)\left(\delta - \frac{\partial f(\cdot)}{\partial x(t)}\right) \tag{1.96}$$

The multipliers $\lambda(t)$ and $\mu(t)$ can be thought of as present-value and current-value shadow prices, respectively, where $\lambda(t)$ is the imputed value of an incremental unit in $x(t)$ from the perspective of $t = 0$, while $\mu(t)$ is the value of an additional unit of $x(t)$ at instant t. For problem (1.90) the complete set of first order conditions, expressed in terms of the current-value Hamiltonian is

$$\begin{aligned} \frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial y_t} &= 0 \\ \dot{\mu} - \delta\mu(t) = -\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial x(t)} &\qquad \dot{x} = \frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial \mu(t)} \\ \mu(T) = F'(\cdot) &\qquad x(0) = a \end{aligned} \tag{1.97}$$

The infinite horizon problem with discounting becomes

$$\begin{aligned} \text{maximize} \int_0^\infty V(\cdot)e^{-\delta t}\,dt \\ \text{subject to} \quad \dot{x} = f(\cdot) \\ x(0) = a \quad \text{given} \end{aligned} \tag{1.98}$$

Assume that a steady state exists and is reachable from $x(0) = a$. The current value Hamiltonian remains unchanged from (1.92). Evaluating Equation (1.95) in steady state implies $\mu = -(\partial V(\cdot)/\partial y)/(\partial f(\cdot)/\partial y)$. Evaluating (1.96) in steady state ($\dot{\mu} = 0$), substituting the expression for μ and isolating δ on the RHS yields:

$$f_x - \frac{V_x}{V_y}f_y = \delta \tag{1.99}$$

which is identical to the steady state expression (1.89) obtained from the analogous discrete-time problem. While discrete- and continuous-time analogues will typically produce identical expressions for steady state they may be subject to different dynamic behavior.

Consider the following problem, which is a simple special case of the above:

$$\begin{aligned} \text{maximize} \int_0^\infty U\big(h(t)\big)e^{-\delta t}\,dt \\ \text{subject to} \quad \dot{x} = F(x) - h(t) \\ x(0) \quad \text{given} \end{aligned}$$

We will assume $U(\cdot)$ to be concave ($U'(\cdot) > 0$, $U''(\cdot) < 0$) and it may be thought of as social utility [or, perhaps, monopoly revenue, in which case $U(h(t)) = p(h(t))h(t)$, where $p(h(t))$ is the inverse demand function for $h(t)$].

The Hamiltonian is

$$\mathcal{H} = U(\cdot)e^{-\delta t} + \lambda(t)[F(\cdot) - h(t)]$$

while the current value Hamiltonian is

$$\tilde{\mathcal{H}} = U(\cdot) + \mu(t)[F(\cdot) - h(t)]$$

where $\mu(t) = \lambda(t)e^{\delta t}$. Using the first two conditions in (1.97)

$$\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial h_t} = U'(\cdot) - \mu(t) = 0$$

$$\dot{\mu} - \delta\mu(t) = -\frac{\partial \tilde{\mathcal{H}}(\cdot)}{\partial x(t)} = -\mu(t)F'(\cdot)$$

Taking the time derivative of the first yields

$$U''(\cdot)\dot{h} - \dot{\mu} = 0$$

Substituting into the second and noting $\mu(t) = U'(\cdot)$ yields

$$\dot{h} = \frac{U'(\cdot)}{U''(\cdot)}[\delta - F'(\cdot)]$$

$\dot{x} = F(x) - h$

which along with the state equation constitutes a coupled nonlinear system of differential equations for the optimal control $h(t)$ and the state variable $x(t)$. This system can be analyzed by standard techniques for *plane* (two-dimensional) *dynamical systems.* Equilibrium occurs where $\dot{h} = \dot{x} = 0$ implying

$$F'(x^*) = \delta$$
$$h^* = F(x^*)$$

where (x^*, h^*) is a steady state equilibrium. Graphically this equilibrium is the intersection of two isoclines $\dot{x} = 0$ and $\dot{h} = 0$ which are drawn in Figure 1.6.

The isoclines divide the positive orthant into four *isosectors* labeled I, II, III, and IV. Each isosector has a *directional* indicating the movement of a point (x, h) over time. For example, a point in isosector I would move in a southwesterly direction under the influence of dynamics which reduce both $x(t)$ and $h(t)$ over time. How did we determine the directionals in isosector I?

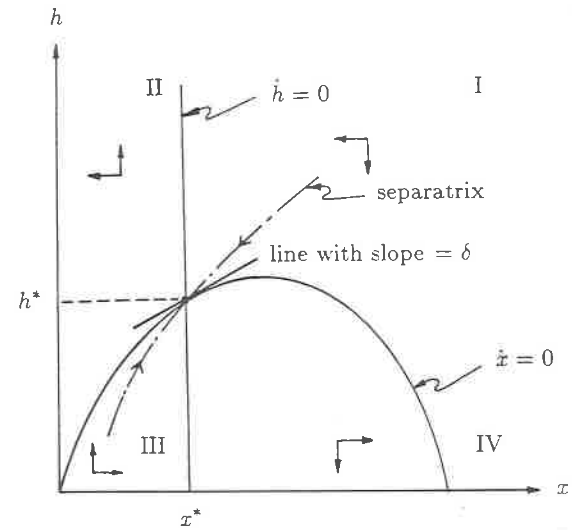


Figure 1.6 Phase plane diagram of the system: $\dot{h} = \frac{U'(\cdot)}{U''(\cdot)}[\delta - F'(x)]$ and $\dot{x} = F(x) - h$ where $\dot{h} = 0$ when $F'(x^*) = \delta$ and $\dot{x} = 0$ when $h = F(x)$.

Along $x = x^*$, $\dot{h} = 0$. If $x > x^*$ the slope of $F(x)$ is less than δ and $[\delta - F'(\cdot)] > 0$. Noting $U'(\cdot)/U''(\cdot) < 0$ (recall our concavity assumption) then for $x > x^*$, $\dot{h} < 0$. For points above the $\dot{x} = 0$ isocline $h > F(x)$, thus $\dot{x} < 0$ and we thus obtain the downward and leftward pointing directional for isosector I. Similar analysis will permit one to determine the directionals in the other sectors.

Isosectors I and III are *convergent* isosectors while isosectors II and IV are *divergent.* The equilibrium (x^*, h^*) would be classified as a *saddle point.* Isosectors I and III each contain a trajectory which converges to (x^*, h^*) and is referred to as a *separatrix.* Taken together the two *separatrices* define the optimal solution trajectories for our infinite horizon problem (since any other trajectory converges either to $h = 0$ or $x = 0$ as $t \to \infty$).[13] If these separatrix curves could actually be computed we would have an explicit optimal harvest policy

$$h^* = h^*(x)$$

specifying the optimal harvest rate corresponding to any given stock level x. Such a rule is called a *feedback or closed-loop control policy,* since h^* is

[13] The convergent separatrices are said to form the *stable manifold.* Isosectors II and IV contain divergent separatrices which form the unstable manifold.

specified as a function of the current state x. A numerical procedure for computing the convergent separatrices will be given in Section 1.6.

In conclusion, let us remark that the situation shown in Figure 1.6 is typical for optimal control problems with a single state variable. With suitable concavity conditions, there exists a unique optimal equilibrium (x^*, h^*), which is a saddle point. The optimal approach to equilibrium follows the separatrix curves which specify a feedback control law.

Many things can go wrong, however, including nonconcavity effects, resulting in the nonexistence of a solution in the classical sense, and multiple equilibria. If the original control problem is time-dependent or if it has two or more state variables, the above phase plane methods do not apply. Normally one must fall back on purely numerical methods, but the subject bristles with difficulties of all kinds (some of the possibilities are discussed, for example, in Clark, 1976).

1.6 Some numerical and graphical techniques

1.6.1 Solution of equations: bisection

In infinite horizon problems we raised the possibility that the solution of the first order conditions might lead to a steady state. When evaluated at steady state the first order conditions reduced to a set of simultaneous equations with the unknowns being the equilibrium state, costate (or Lagrange), and control variables. By elimination, this set of equations might be further reduced to a single equation defining the optimal steady state value for a particular variable [see the discussion of Equation (1.89)].

Sometimes the steady state equation may be explicitly solved by algebra. In other instances it might only be written implicitly as, say, $g(x) = 0$. A steady state value, x^*, is thus a "zero" or root of $g(x)$, i.e., $g(x^*) = 0$. We will discuss two numerical techniques for solving for the zero of an equation. The first technique is interval bisection.

First we must "bracket" the desired solution $x^* \in (x_1, x_2)$ by determining two values x_1 and x_2 for which $g(x_1) < 0$ and $g(x_2) > 0$ (or vice versa). This situation is shown in Figure 1.7.

Define $z = (x_1 + x_2)/2$. Then we may approximate x^* by employing the following simple algorithm:

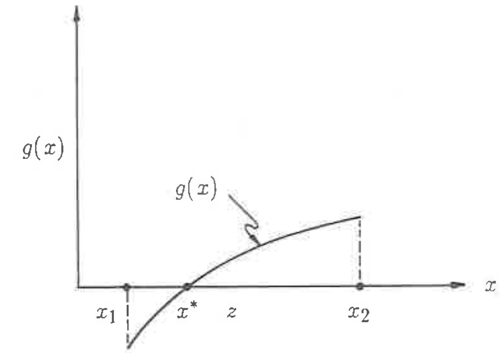


Figure 1.7 Interval bisection.

(a) $z = (x_1 + x_2)/2$
(b) if $|g(z)| \le \epsilon$, stop (you've found x^* within some preassigned tolerance $\epsilon > 0$)
(c) if $|g(z)| > \epsilon$, then examine $g(z)$
(d) if $g(z) > 0$, redefine $x_2 = z$ otherwise $x_1 = z$ and return to (a).

It is clear from Figure 1.7 that the interval (x_1, x_2) will "zero in" on the solution x^*. The speed of convergence can be estimated by noting that each loop cuts the interval in half. Thus 10 loops will increase the accuracy by $(1/2)^{10} \approx 0.001$, or 3 decimals.

Program 1.2 solves $g(x) = x^3 + x + 1 = 0$. We start by trial and error until we find suitable x_1 and x_2. In our case $x_1 = -1$ and $x_2 = 0$, since $g(-1) = -1$ and $g(0) = +1$. The solution is $x^* = -0.682328$.

The second technique for finding a zero of $g(x)$, called Newton's method, is useful provided the first derivative, $g'(x)$, can be obtained.

1.6.2 Newton's method

If x^* is a zero of $g(x)$ then it is also a "fixed-point" of the function

$$f(x) = x + h(x)g(x) \tag{1.100}$$

in the sense that $f(x^*) = x^*$ since $g(x^*) = 0$. Here $h(x)$ is an arbitrary function. A zero of $g(x)$ may be found as the limit of an iteration sequence $\{x_t\}$ if the initial guess, x_0, is "close" to the zero, and if $f(x)$ is chosen so that $|f'(x^*)| < 1$. Convergence will be most rapid if we define $f(x)$ so that $f'(x^*) = 0$. Taking the derivative of (1.100) and setting it equal to

```
10 REM PROGRAM 1.2: INTERVAL BISECTION
20 EPS = .000001
30 X1 = -1: X2 = 0
40 LPRINT "    Z           G(Z)"
50 LPRINT "---------------------------"
60 Z = .5 * (X1 + X2)
70 G = Z^3 + Z + 1
80 LPRINT Z,G
90 IF G > 0 THEN X2 = Z ELSE X1 = Z
100 IF ABS(G) > EPS THEN GOTO 60
110 END
```

```
    Z             G(Z)
----------------------------
-.5              .375
-.75            -.171875
-.625            .1308594
-.6875          -1.245117E-02
-.65625          6.112671E-02
-.671875         2.482987E-02
-.6796875        6.313801E-03
-.6835938       -3.037453E-03
-.6816406        1.646042E-03
-.6826172       -6.937981E-04
-.682129         4.765988E-04
-.6823731       -1.084805E-04
-.682251         1.841188E-04
-.682312         3.784895E-05
-.6823425       -3.528595E-05
-.6823273        1.311302E-06
-.6823349       -1.692772E-05
-.6823311       -7.867813E-06
-.6823292       -3.33786E-06
-.6823283       -9.536743E-07
```

Program 1.2 Interval bisection.

zero yields

$$f'(x) = 1 + h'(x)g(x) + h(x)g'(x) = 0 \tag{1.101}$$

Since $g(x^*) = 0$ we obtain $f'(x^*) = 0$ if

$$h(x^*) = -1/g'(x^*) \tag{1.102}$$

provided $g'(x^*) \neq 0$. The above result leads to the iteration equation

$$x_{t+1} = f(x_t) = x_t - g(x_t)/g'(x_t) \tag{1.103}$$

Starting from some x_0 sufficiently close to x^* Equation (1.103) will ultimately lead to an $x_{t+1} \approx x^*$, with $|g(x_{t+1})| < \epsilon$ for any prescribed "tolerance" $\epsilon > 0$ (the tolerance can be included in the iteration program).

The Newton method has a simple graphical interpretation. Equation (1.103) implies

$$g'(x_t) = -g(x_t)/(x_{t+1} - x_t) \tag{1.104}$$

and the sequence of convergence is shown in Figure 1.8.

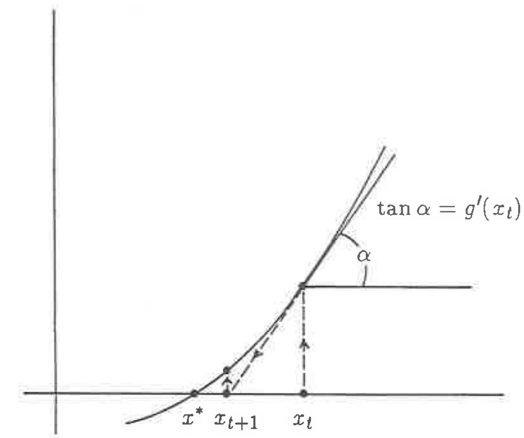


Figure 1.8 Graphical representation of Newton's method.

```
10 REM PROGRAM 1.3: NEWTON'S METHOD
20 LPRINT "  X                 G(X)":LPRINT "
30 X = -.5
40 EPS = .000001
50 G = X^3 + X + 1
60 LPRINT X,G
70 IF ABS(G) < EPS GOTO 110
80 DG = 3 * X^2 + 1
90 X = X - G / DG
100 GOTO 40
110 END
```

```
  X                G(X)

-.5               .375
-.7142858        -7.871723E-02
-.6831798        -2.043366E-03
-.6823284        -1.430512E-06
-.6823278         0
```

Program 1.3 Newton's method

As a simple example, let us again solve the equation

$$g(x) = x^3 + x + 1 = 0$$

Recall that $g(0) = 1$ and $g(-1) = -1$ and that from the method of interval bisection we obtained $x^* = -0.682328$. The first derivative is

```
10 REM PROGRAM 1.4: NEWTON'S METHOD
20 LPRINT "  X                    G(X)":LPRINT "
30 X=3
40 EPS = .000001
50 G = (X-1) * EXP(-X)
60 LPRINT X,G
70 IF ABS(G) < EPS GOTO 110
80 DG = (2-X) * EXP(-X)
90 X = X - G / DG
100 GOTO 40
110 END
```

```
 X              G(X)

3              9.957415E-02
5              2.695179E-02
6.333334       9.472552E-03
7.564103       3.405081E-03
8.743826       1.234697E-03
9.89211        4.496931E-04
11.01882       1.642118E-04
12.1297        6.006517E-05
13.22842       2.19959E-05
14.31748       8.061595E-06
15.39866       2.956432E-06
16.4733        1.084728E-06
17.54239       3.981398E-07
```

Program 1.4 Newton's method

$g'(x) = 3x^2 + 1$. The sequence $\{x_t\}$, starting, say, with $x_0 = -0.5$ can be easily obtained from a pocket calculator or a microcomputer. A BASIC program and its output are listed in Program 1.3; note the more rapid convergence to the solution $x^* = -0.682328$.

The speed of Newton's algorithm can be estimated from the fact (which can be proved) that Newton's method improves accuracy by about 2 decimals for each iteration. Thus Newton's method will take about 3 (or 4) steps for 6-figure accuracy, compared to about 20 steps for the bisection algorithm. Which method to use depends primarily on how hard it is to work out the derivative $g'(x)$. In some of the problems given later you'll probably be better off with bisection!

Newton's method can be extended to the solution of simultaneous equations (where it is usually called the Newton–Raphson method). Most computer centers have library routines using this method. The only difficulty will usually be in finding a good first approximation. An indication of what might go wrong is the following attempt to solve $g(x) = (x-1)e^{-x} = 0$ with

$x_0 = 3$, using Program 1.4. The only solution is $x = 1$. What happened? (Make a sketch.)

1.6.3 Eigenvalues

Once a steady state solution to a dynamic optimization model has been identified one would typically want to determine its "local" stability, that is, the dynamic behavior of the system near the steady state equilibrium. For nonlinear plane autonomous systems we could proceed by drawing the isoclines and identifying the directionals in the various isosectors as was done in Figure 1.6 for the nonlinear renewable resource problem posed at the end of the previous section. For numerical problems the local stability can be determined by calculating the characteristic roots or eigenvalues of the "linearized" dynamical system evaluated at steady state. Specifically, suppose the maximum principle, after a certain amount of manipulation, resulted in the following dynamical system for the state variable $x(t)$ and control variable $y(t)$

$$\left.\begin{aligned} \dot{x} &= F(x,y) \\ \dot{y} &= G(x,y) \end{aligned}\right\} \tag{1.105}$$

Let (x^*, y^*) be an equilibrium, so that $F(x^*, y^*) = G(x^*, y^*) = 0$. Then if $F(\cdot)$ and $G(\cdot)$ are smooth we can approximate the system with

$$\begin{aligned} \dot{x} &= F_x(\cdot)(x - x^*) + F_y(\cdot)(y - y^*) + 0[(x - x^*)^2 + (y - y^*)^2] \\ \dot{y} &= G_x(\cdot)(x - x^*) + G_y(\cdot)(y - y^*) + 0[(x - x^*)^2 + (y - y^*)^2] \end{aligned} \tag{1.106}$$

where the partials of $F(\cdot)$ and $G(\cdot)$ are evaluated at (x^*, y^*) and where $0\{\cdot\}$ are terms of order two.

It can be shown that if the linear system obtained by omitting the higher order terms in (1.106) is structurally stable (node, saddle point, or a spiral) then the nonlinear system (1.105) has the same behavior in the neighborhood of (x^*, y^*) (see Birkhoff and Rota, 1969, p. 135, for the proof).

Let $a_{11} = F_x(x^*, y^*)$, $a_{12} = F_y(x^*, y^*)$, $a_{21} = G_x(x^*, y^*)$ and $a_{22} = G_y(x^*, y^*)$. Then the eigenvalues of the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{1.107}$$

are given by the characteristic equation[14]

$$\det[A - RI] = \begin{vmatrix} a_{11} - R & a_{12} \\ a_{21} & a_{22} - R \end{vmatrix} = R^2 - (a_{11} + a_{22})R + (a_{11}a_{22} - a_{21}a_{12}) = 0 \tag{1.108}$$

implying

$$R_1 = \frac{1}{2}\left(s + \sqrt{s^2 - 4d}\right) \qquad R_2 = \frac{1}{2}\left(s - \sqrt{s^2 - 4d}\right) \tag{1.109}$$

where $s = (a_{11} + a_{22})$ and $d = (a_{11}a_{22} - a_{21}a_{12})$. R_1 and R_2 will be real and distinct (when $(s^2 - 4d) > 0$) or complex conjugates (when $(s^2 - 4d) < 0$) and the equilibrium (x^*, y^*) of the nonlinear system (1.105) may be locally classified as

(a) an unstable node if $R_1, R_2 > 0$
(b) a stable node if $R_1, R_2 < 0$
(c) a saddle point if $R_1 < 0 < R_2$ or $R_2 < 0 < R_1$
(d) an unstable spiral if R_1, R_2 are complex with positive real part
(e) a stable spiral if R_1, R_2 are complex with negative real part

Complex if $s^2 - 4d < 0$, implying the square root of a negative number

Consider the following problem:

$$\begin{aligned} \text{maximize} \quad & \int_0^\infty [20 \ln x(t) - 0.10 y(t)^2]\, dt \\ \text{subject to} \quad & \dot{x} = y(t) - 0.10x(t) \\ & x(0) = 10 \end{aligned}$$

The Hamiltonian for this problem is

$$\mathcal{H}(\cdot) = 20 \ln x(t) - 0.10y(t)^2 + \lambda(t)[y(t) - 0.10x(t)]$$

[14] The characteristic equation arises from expansion of the determinant

$$\det(A - RI) = \begin{vmatrix} a_{11} - R & a_{12} \\ a_{21} & a_{22} - R \end{vmatrix} = 0$$

with first order necessary conditions requiring

$$\begin{aligned} \frac{\partial \mathcal{H}(\cdot)}{\partial y(t)} &= -0.20y(t) + \lambda(t) = 0 \\ \dot{\lambda} &= -\frac{\partial \mathcal{H}(\cdot)}{\partial x(t)} = -\frac{20}{x(t)} + 0.10\lambda(t) \\ \dot{x} &= \frac{\partial \mathcal{H}(\cdot)}{\partial \lambda(t)} = y(t) - 0.10x(t) \end{aligned}$$

In steady state $\dot{x} = \dot{\lambda} = 0$, implying

$$\begin{aligned} \lambda &= 0.20y \\ 0 &= -20/x + 0.10\lambda \\ y &= 0.10x \end{aligned}$$

The steady state equations are easily solved yielding $x = 100$, $y = 10$, $\lambda = 2$. What is the character of the equilibrium point? To evaluate the stationary state $(100, 10)$ in (x, y)-space we first solve for an expression for $\dot{y}$. Taking the time derivative of the maximum condition implies $\dot{\lambda} = 0.20\dot{y}$. Substituting into the second condition implies $0.20\dot{y} = -20/x(t) + 0.02y(t)$. Solving for $\dot{y}$ the first order conditions imply:

$$\begin{aligned} \dot{x} &= -0.10x(t) + y(t) = F(x(t), y(t)) \\ \dot{y} &= -100/x(t) + 0.10y(t) = G(x(t), y(t)) \end{aligned}$$

The isoclines for this system are obtained by setting $\dot{x} = \dot{y} = 0$ implying

$$\begin{aligned} y &= 0.10x \quad \text{when} \quad \dot{x} = 0 \\ y &= 1{,}000/x \quad \text{when} \quad \dot{y} = 0 \end{aligned}$$

The isoclines are drawn in the (x, y) phase plane and are shown in Figure 1.9. Note that the isoclines intersect at the steady state $(100, 10)$. The isoclines define four isosectors indicated by roman numerals. The directionals can be verified in a manner similar to that employed in the discussion of Figure 1.6. The linearized matrix is

$$A = \begin{pmatrix} -0.10 & 1 \\ 0.01 & 0.10 \end{pmatrix}$$

which leads to calculations of $s = 0$, $d = -0.02$ and eigenvalues of $R_1 = \sqrt{0.02}$, $R_2 = -\sqrt{0.02}$ confirming our qualitative analysis that $(100, 10)$ is a saddle point.

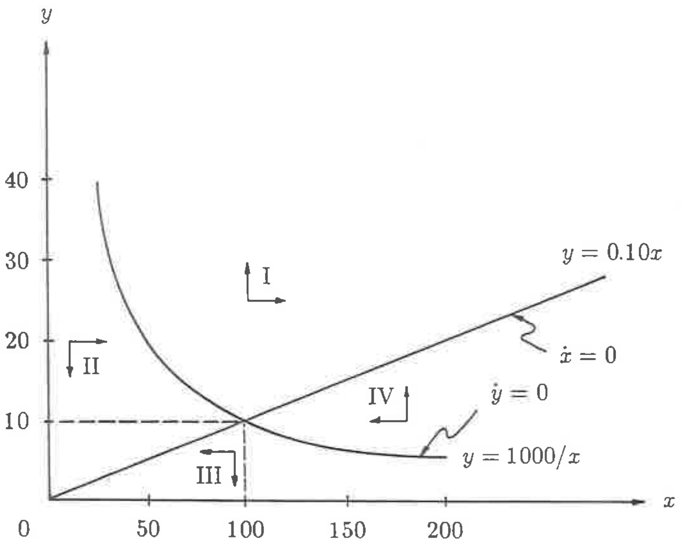


Figure 1.9 A phase plane analysis of the system $\dot{x} = -0.10x(t) + y(t)$, $\dot{y} = -100/x(t) + 0.1y(t)$.

1.6.4 Eigenvectors

Having solved for the eigenvalues of a matrix one can also solve for the associated *eigenvectors*. The eigenvectors are required if one wishes to diagonalize a symmetric matrix. Also, as we shall see, they can be used in solving for the saddle point separatrices in a two-dimensional nonlinear dynamical system.

By definition, R is an eigenvalue of the matrix A if

$$\det(A - RI) = 0$$

This means that the matrix $(A - RI)$ is "singular," i.e., that the linear system of equations

$$(A - RI)X = 0$$

has a nonzero solution vector X. Any such vector is called an *eigenvector* associated with R.

In the previous subsection we worked through an example where

$$A = \begin{pmatrix} -0.10 & 1 \\ 0.01 & 0.10 \end{pmatrix}$$

with eigenvalues $R_1 = \sqrt{0.02}$ and $R_2 = -\sqrt{0.02}$. Thus the eigenvector system for $R = R_1$ is

$$(A - R_1 I)X = \begin{pmatrix} -0.2414 & 1 \\ 0.01 & -0.0414 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The two rows of the matrix $(A - R_1 I)$ are *linearly dependent*. If x, y satisfy the single equation

$$-0.2414x + y = 0$$

then (x, y) will be an eigenvector. For example

$$V_1 = (x, y) = (1, 0.2414)$$

is one choice; any other eigenvector is a scalar multiple of V_1. By a similar calculation we find that

$$V_2 = (1, 0.0414)$$

is an eigenvector corresponding to the eigenvalue $R_2 = -\sqrt{0.02}$.

The problem of determining the eigenvalues and eigenvectors of a matrix arises in many areas of applied mathematics. Efficient computer algorithms are available for this problem; an excellent textbook is Strang (1980).

1.6.5 Numerical solution of differential equations

Frequently it is not possible to obtain closed form solutions for differential equation systems such as those obtained via the maximum principle. In numerical problems the computer can be used to obtain accurate approximate numerical solutions to such systems. Of the many known algorithms, we will discuss here only the popular Runge–Kutta method.

We consider an initial value problem

$$\frac{dy}{dt} = F(t, y) \qquad t_0 \le t \le t_1$$
$$y(t_0) = y_0 \quad \text{given} \tag{1.110}$$

Here $y = (y^1, \ldots, y^N)$ in general denotes a vector of N unknown functions $y^i(t)$. We consider a uniform mesh on the interval $[t_0, t_1]$ given by

$$t_j = t_0 + jh \qquad h = \frac{t_1 - t_0}{n} \qquad j = 0, 1, \ldots, n$$

We wish to obtain approximate values for the solution $y(t)$ at each mesh point t_j. If n is sufficiently large (i.e., h sufficiently small), interpolation of these values will closely approximate the desired solution over $t_0 \le t \le t_1$.

The Runge–Kutta algorithm is given by

$$y_0 \quad \text{given}$$

$$y_{j+1} = y_j + \frac{h}{6}(k_{j1} + 2k_{j2} + 2k_{j3} + k_{j4}) \tag{1.111}$$

where

$$\begin{aligned} k_{j1} &= F(t_j, y_j) \\ k_{j2} &= F(t_j + h/2, y_j + hk_{j1}/2) \\ k_{j3} &= F(t_j + h/2, y_j + hk_{j2}/2) \\ k_{j4} &= F(t_j + h, y_j + hk_{j3}) \end{aligned} \tag{1.112}$$

This algorithm is easily coded for the computer (see Program 1.5).

As a first example, consider the logistic equation

$$\dot{y} = F(y) = ry(1 - y/K) \qquad 0 \le t \le 100 \tag{1.113}$$

where now y is one-dimensional. We take

$$y_0 = 10 \quad r = 0.10 \quad K = 200$$

The output of Program 1.5 compares the exact solution (Clark, 1976, p. 11)

$$y(t) = \frac{K}{1 + ce^{-rt}} \quad \text{where} \quad c = 19$$

with the values computed via Runge–Kutta, with $h = 10$ ($n = 10$), and also via the simplistic difference scheme

$$y_{j+1} = y_j + F(y_j)$$

Note that, in spite of the coarse mesh ($h = 10$), even the difference scheme is not bad (maximum relative error of about 6%), while the Runge–Kutta method is quite accurate (maximum error .006%). Of course, the real usefulness of Runge–Kutta (and other numerical methods) comes when no analytic solution is possible.


```
10 REM PROGRAM 1.5: RUNGE-KUTTA EXAMPLE
20 DIM YRK(100),YDIFF(100)
30 DATA 0.10,200,10
40 READ R,K,YRK(0)
50 C = K / YRK(0) - 1
60 YDIFF(0) = YRK(0) - 1
70 LPRINT" T              YEXACT(T)      YRK(T)          YDIFF(T)"
80 LPRINT
90 DEF FN F(X) = R * X * (1-X/K)
100 FOR T = 0 TO 99
110   Y = YRK(T)
120 K1 = FN F(Y)
130   K2 = FN F(Y + K1/2)
140   K3 = FN F(Y + K2/2)
150   K4 = FN F(Y + K3)
160   YRK(T+1) = Y + (K1 + 2*K2 + 2*K3 + K4) / 6
170   Y1 = YDIFF(T)
180   YDIFF(T+1) = Y1 + FN F(Y1)
190 NEXT
200 FOR T = 0 TO 100 STEP 10
210    YEXACT = K / (1 + C * EXP(-R*T))
220    LPRINT T,YEXACT,YRK(T),YDIFF(T)
230 NEXT
240 END
```

T	YEXACT(T)	YRK(T)	YDIFF(T)
0	10	10	9
10	25.0322	25.03219	21.90572
20	56.00091	56.00088	48.9288
30	102.7773	102.7773	92.88643
40	148.3683	148.3682	140.7811
50	177.3017	177.3016	173.8964
60	191.0044	191.0044	189.9599
70	196.5939	196.5939	196.3678
80	198.7333	198.7333	198.7167
90	199.5321	199.5321	199.5505
100	199.8276	199.8276	199.843

Program 1.5 Runge–Kutta example.

As a two-dimensional example, let us consider the classic Lotka–Volterra predator-prey model (Clark, 1976, p. 194):

$$\begin{aligned} \dot{x} &= rx - \alpha xy \\ \dot{y} &= \beta xy - sy \end{aligned} \tag{1.114}$$

This has an equilibrium at the point

$$\bar{x} = \frac{s}{\beta} \qquad \bar{y} = \frac{r}{\alpha}$$

This point is a center of the corresponding linearized system, and it can

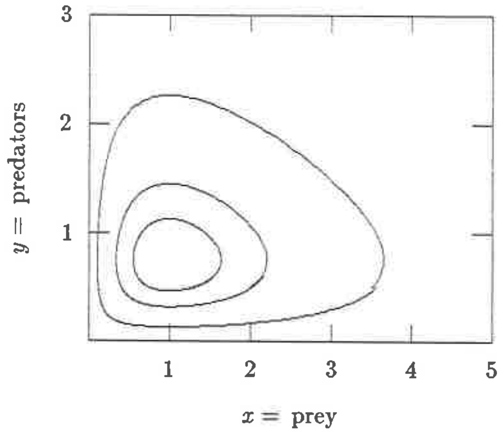


Figure 1.10 Lotka–Volterra curves plotted by microcomputer, using the Runge–Kutta method.

be shown that the trajectories of (1.114) consist of closed orbits circling counterclockwise around the equilibrium (Clark, ibid).

Figure 1.10 shows three such orbits computed via the Runge-Kutta scheme (see Program 1.6) using the parameter values

$$r = 1 \qquad s = 1 \qquad \alpha = 2 \qquad \beta = 1$$

(The listed program can be used to plot phase plane trajectories for general systems.)

For further discussion of numerical solution methods, see, e.g., Henrici (1982).

1.6.6 Computation of separatrices

Toward the end of Section 1.5 we posed the nonlinear renewable resource problem

$$\begin{aligned} &\text{maximize} && \int_0^\infty U(h(t))e^{-\delta t}\,dt \\ &\text{subject to} && \dot{x} = F(x) - h(t) \\ & && x(0) \quad \text{given} \end{aligned}$$

whose analysis led to the phase plane diagram in Figure 1.6. In isosectors I and III of that diagram we identified two separatrices which defined the

```
10 REM PROGRAM 1.6: PHASE PLANE PLOTTER
20 REM  PLOTS PHASE PLANE DIAGRAM FOR THE FOLLOWING FUNCTIONS
30 DEF FN F(X,Y) = X - 2 * X * Y
40 DEF FN G(X,Y) = X * Y - Y
50 INPUT "HORIZONTAL RANGE";X1,X2
60 INPUT "VERTICAL RANGE";Y1,Y2
70 INPUT "HORIZONTAL BLIP SCALE";XB
80 INPUT "VERTICAL BLIP SCALE";YB
90 INPUT "NUMBER OF CURVES (MAX 3)";NCURVES
100 DIM X0(3),Y0(3),TF(3),DT(3)
110 FOR N = 1 TO NCURVES
120   PRINT "STARTING POINT FOR CURVE " N
130   INPUT X0(N),Y0(N)
140   INPUT "TIME LIMIT"; TF(N)
150   INPUT "MESH";DT(N)
160 NEXT
170 '  ***** DRAW WINDOW *****
180 SCREEN 2: CLS: KEY OFF: SCALE = 1.5
190 WINDOW (X1,Y1)-(SCALE*X2,SCALE*Y2)
200 LINE (X1,Y1)-(X2,Y2),,B
210 XL = (X2-X1) * .05: YL = (Y2-Y1) * .05
220 X = X1
230 WHILE X <= X2-XB
240  Y = Y1: X = X + XB: LINE (X,Y)-(X,Y+YL)
250  Y = Y2 - YL: LINE (X,Y)-(X,Y+YL)
260 WEND
270 Y = Y1
280 WHILE Y <= Y2 - YB
290  X = X1: Y = Y + YB: LINE (X,Y)-(X+XL,Y)
300  X = X2 - XL: LINE (X,Y)-(X+XL,Y)
310 WEND
320 '  ***** START GRAPHING  *****
330 FOR NC = 1 TO NCURVES
340 N = TF(NC) / DT(NC)
350 X = X0(NC): Y= Y0(NC): H = DT(NC)
360 LINE (X,Y)-(X,Y)
370 FOR I = 1 TO N
380   KX1 = FN F(X,Y)
390   KY1 = FN G(X,Y)
400   XT = X + H * KX1 / 2: YT = T + H * KY1 / 2
410   KX2 = FN F(XT,YT)
420   KY2 = FN G(XT,YT)
430   XT = X + H * KX2 / 2: YT = Y + H * KY2 / 2
440   KX3 = FN F(XT,YT)
450   KY3 = FN G(XT,YT)
460   XT = X + H * KX3: YT = Y + H * DY3
470   KX4 = FN F(XT,YT)
480   KY4 = FN G(XT,YT)
490   X = X + (KX1 + 2 * (KX2 + KX3) + KX4) * H / 6
500   Y = Y + (KY1 + 2 * (KY2 + KY3) + KY4) * H / 6
510   LINE -(X,Y)
520 NEXT I
530 NEXT NC
540 DEF SEG=&HB000
550 BSAVE "PROG1-5.PIC",0,&HB000
560 END
```

Program 1.6 Phase plane plotter

optimal feedback control $h^* = h^*(x)$. A similar result holds for the example at the end of the last section. This subsection discusses how separatrices might be computed and graphed using a microcomputer.

The numerical solution of ordinary differential equations is simple and straightforward. The convergent separatrices, however, pose a slight problem since one does not know, a priori, where they "begin." An efficient approach is first to approximate the separatices in the neighborhood of the equilibrium point (x^*, y^*) by linearizing

$$\begin{aligned} \dot{x} &= F(x) - h(t) \\ \dot{h} &= \frac{U'(\cdot)}{U''(\cdot)}[\delta - F'(x)] \end{aligned} \tag{1.115}$$

about that point.

To be explicit, we first expand the functions on the right side of (1.115) using Taylor's expansion:

$$\begin{aligned} G(x,h) = G(x^*, h^*) &+ G_x(x^*, h^*)(x - x^*) \\ &+ G_h(x^*, h^*)(h - h^*) + \text{smaller order terms} \end{aligned}$$

After a simple calculation, we obtain the linearized system (with $\xi = x - x^*$, $\eta = h - h^*$):

$$\left.\begin{aligned} \dot{\xi} &= \delta\xi - \eta \\ \dot{\eta} &= c\xi \end{aligned}\right\} \tag{1.116}$$

where $c = -U'(h^*)F''(x^*)/U''(h^*) < 0$.

The linear system can be solved by elementary methods. First, the eigenvalues R are the solutions of

$$\det\begin{pmatrix} \delta - R & -1 \\ c & -R \end{pmatrix} = R^2 - \delta R + c = 0$$

so that

$$R_1 = \frac{1}{2}(\delta + \sqrt{\delta^2 - 4c}) \quad R_2 = \frac{1}{2}(\delta - \sqrt{\delta^2 - 4c})$$

Since $c < 0$, these eigenvalues are real and of opposite sign.[15] Hence (as we already know), the equilibrium point is a saddle point.

[15] It is of interest to note here that if $c > 0$ (as could happen if either U or F failed to be concave), then not only does the saddle point geometry disappear, but in fact the maximum principle does not work! Indeed, no optimal control even exists under these conditions.

Let V_1, V_2 denote the eigenvectors corresponding to R_1, R_2, respectively. The general solution of the linear system (1.116) is then

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = aV_1 e^{R_1 t} + bV_2 e^{R_2 t} \qquad (a, b \text{ constant})$$

Since $R_2 < 0 < R_1$, the only trajectories which converge to the origin as $t \to +\infty$ are those for which $a = 0$. Thus the eigenvectors $\pm V_2$ determine the directions of the convergent separatrices. (A numerical example is worked out below.)

Returning now to the (x,h)-plane, we have shown that the eigenvectors $\pm V_2$ (with $R_2 < 0$) determine the directions along which the separatrices approach the equilibrium (x^*, h^*). The actual (curved) separatrices themselves may now be computed by numerical integration of the system (1.115). The best way to do this is to start with a point (x_0, h_0) near (x^*, h^*) and on the linearized separatrix found above; then solve (1.115) with time reversed (i.e., change the signs of the right sides).

Example. Assume that

$$\begin{aligned} F(x) &= x(1-x) \\ U(h) &= h(1-h) \\ \delta &= 0.1 \end{aligned}$$

Then $F'(x) = 1 - 2x$, so that $\delta - F'(x^*) = 0$ implies $x^* = 0.45$ and $h^* = F(x^*) = 0.248$. [For the monopolist price $= p(h) = U(h)/h$ we have $p(0) = 1$ and $p(h^*) = 0.752$, i.e., the price at optimal sustained yield is only 75% of the choke-off price $p(0)$.]

The plane system (1.115) is

$$\begin{aligned} \dot{x} &= x(1-x) - h \\ \dot{h} &= -(1-2h)(x - 1.45) \end{aligned}$$

and the linearized version is

$$\begin{aligned} \dot{\xi} &= 0.1\xi - \eta \\ \dot{\eta} &= -0.504\xi \end{aligned}$$

The eigenvalues of the matrix

$$\begin{pmatrix} 0.1 & -1 \\ -0.504 & 0 \end{pmatrix}$$

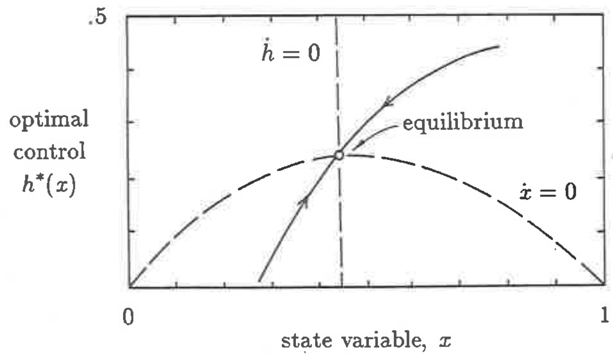


Figure 1.11 Convergent separatrices, defining optimal feedback policy $h^* = h^*(x)$, as obtained by numerical computation. The point (x^*, h^*) moves along the separatrix, from (x_0, h_0) to (x^*, h^*) as $t \to +\infty$.

are the solutions of the equation $R^2 - 0.1R - 0.504 = 0$, which are

$$R_1 = 0.762 \qquad \text{and} \qquad R_2 = -0.662$$

with corresponding eigenvectors

$$V_1 = (1, -0.662) \qquad \text{and} \qquad V_2 = (1, 0.762)$$

Thus the convergent separatrices have slope 0.762 at the equilibrium point (x^*, h^*). The resulting nonlinear separatrices (shown in Figure 1.11) were calculated in a few seconds using the Runge–Kutta algorithm.

1.7 Problems

P1.7.1. Consider the competitive (price-taking) firm which seeks to maximize profits subject to the following multiple-output production function:

$$\phi(Y_1, Y_2; X_1, X_2) \equiv 2Y_1^2 + Y_2^2 - 100X_1^{0.25}X_2^{0.75} \equiv 0$$

If $P_1 = 5$ is the per unit price for output Y_1, $P_2 = 10$ is the per unit price for output Y_2, $R_1 = 1$ is the per unit cost for input X_1, and $R_2 = 2$ is the per unit cost for input X_2, solve for the optimal values of Y_1, Y_2, X_1, and X_2. Note: Profit may be written as:

$$\pi = \sum_{i=1}^{2} P_i Y_i - \sum_{j=1}^{2} R_j X_j$$

Answer. $Y_1 = 42.37$, $Y_2 = 169.49$, $X_1 = 238.44$, $X_2 = 357.66$, $\lambda = 0.0295$.

P1.7.2. You have been appointed manager of a project with the objective of bringing 1,000 hectares under irrigation within 10 years ($t = 0, 1, \ldots, 9$). At the present time there is no irrigated land in the project area. Let X_t represent the number of irrigated hectares in year t. Thus,

$$X_{t+1} = X_t + U_t$$

where U_t is the number of additional hectares irrigated in year t, and $X_0 = 0$, $X_{10} = 1{,}000$ are given.

You are asked to *minimize* the present value of project costs where you have determined that

$$C_t = cU_t^2 \qquad c = \text{constant}$$

is the cost of irrigating U_t additional hectares in period t. You are told that $c = 1$ and $\rho = 1/(1+\delta)$, where $\delta = 0.10$, is the appropriate discount factor. Solve for the optimal time path for incremental and cumulative hectares under irrigation. Explain.

Answer

t	0	1	2	3	4	5	6	7	8	9	10
U_t	62.75	69.02	75.92	83.51	91.87	101.05	111.16	122.27	134.50	147.95	–
X_t	0	62.75	131.77	207.69	291.20	383.07	484.12	595.28	717.55	852.05	1000

Discounting implies that irrigation costs should be delayed to future periods, so the amount of irrigation increases with t. Also marginal costs discounted to $t = 0$ are all equal.

P1.7.3. Consider the following problem

$$\text{maximize} \quad J = \int_0^1 \{10q(t) - 0.01q(t)^2\} e^{-0.20t}\, dt$$

$$\text{subject to} \quad \dot{X}(t) = -q(t)$$

$$X(0) = 1 \qquad X(1) = 0$$

Solve for the optimal time path $X^*(t)$.

Answer. $X^*(t) = 2253.8e^{0.2t} - 500t - 2252.8$

P1.7.4. Consider the continuous-time version of the irrigation problem (Problem 1.7.2) which might be stated

$$\text{minimize} \quad C = \int_0^{10} U(t)^2 e^{-0.10t}\, dt$$

$$\text{subject to} \quad \dot{X}(t) = U(t)$$

$$X(0) = 0 \qquad X(10) = 1{,}000$$

Solve for the equation defining the optimal time path $X^*(t)$. Compare the values of $X^*(t)$ and X_t^* at $t = 3, 5, 7, 9$. Why is there a slight difference?

Answer. $X(t) = 581.98\,(e^{0.10t} - 1)$

t	3	5	7	9
$X(t)$	203.61	377.54	589.98	849.46
X_t	207.69	383.07	595.28	852.05

Continuous discounting at 0.10 results in an *effective* annual discount rate that is greater than an annual discount rate of 0.10. If i is the annual (discrete-time) discount rate and δ is to be the corresponding continuous (or instantaneous) discount rate, then $\delta = \ln(1 + i)$. An $i = 0.10$ in Problem 1.7.2 would correspond to a $\delta = 0.0953$ in this problem. By solving the problem for $\delta = 0.10$ you have increased the effective discount rate and it is optimal to postpone some irrigation until just prior to $t = 10$. (If you solve it for $\delta = 0.0953$ you will get the same results as in Problem 1.7.2.)

P1.7.5. The rate at which natural gas is pumped from a deposit is given by

$$\dot{X} = -aXE$$

where $X(t)$ = amount of gas remaining in the deposit (million cu. ft.—MCF), $E(t)$ = input of pumping energy (kilowatts), and a = constant.

Let p denote the well-head price of gas and let $c(E)$ denote the cost of pumping. Given $a = .01$ /kw year, find the energy input $E(t)$ that

$$\text{maximizes} \quad J = \int_0^1 [paX(t)E(t) - c(E(t))]\, dt$$

$$\text{subject to} \quad \dot{X} = -aXE$$

$$X(0) = 1{,}000 \qquad x(1) = 500$$

(First use the maximum principle to show that $\dot{E} \equiv 0$.)

Given $p = \$2{,}000$ /MCF and $c(E) = 100E^2$ (\$/yr), calculate the maximized profit objective J^*. Also determine explicitly the shadow price $\lambda(t)$. Why does $\lambda(t)$ decrease?

Answer. $E^* = 69.3$ kw, $J^* = \$520{,}000$, $\lambda(t) = \$(2{,}000 - 1{,}386e^{0.693t})$.

The shadow price represents the marginal value of gas in the deposit, and this decreases as the deposit is exploited.

(This example can be generalized to cover other aspects of exhaustible resource economics—see Chapter 3.)

P1.7.6. Assume the dynamics of two competing species may be characterized by the following dynamical system:

$$\dot{X}(t) = rX(t)(1 - X(t)/K) - \alpha X(t)Y(t)$$

$$\dot{Y}(t) = sY(t)(1 - Y(t)/L) - \beta X(t)Y(t)$$

Draw the isoclines, determine the directionals in each isosector and calculate the characteristic roots at the equilibrium when

(a) $r = 0.8$, $K = 200$, $\alpha = 0.002$
$s = 0.5$, $L = 100$, $\beta = 0.001$

(b) As in part (a) except $\alpha = 0.02$, $\beta = 0.01$

(c) As in part (a) except $\beta = 0.01$

Answer

(a) Equilibrium at $X = 166.67$, $Y = 66.67$. Characteristic roots are $R_1 = -0.28$, $R_2 = -0.72$. The equilibrium is a stable node.

(b) There are actually *three* equilibria in this case
   (i) At $X = 33.33$, $Y = 33.33$ we obtain $R_1 = 0.32$, $R_2 = -0.62$, a saddle point.
   (ii) At $X = 200$, $Y = 0$ we obtain $R_1 = -0.80$, $R_2 = -1.5$, a stable node, and
   (iii) At $X = 0$, $Y = 100$ we obtain $R_1 = -0.50$, $R_2 = -1.2$, also a stable node.

   Whether you end up at (33.33, 33.33), (200,0) or (0,100) depends on the initial conditions $(X(0), Y(0))$.

(c) Equilibrium at $X = 200$, $Y = 0$, same analysis as (b) (ii) above.

P1.7.7. Consider the following problem

$$\text{maximize} \quad \sum_{t=0}^{\infty} \rho^t Y_t$$

$$\text{subject to} \quad X_{t+1} = 100X_t(1 + 0.18X_t)^{-1} - Y_t$$

$$X_0 = 550 \qquad 0 \le Y_t \le 453$$

For $\rho = 1/(1+\delta)$, $\delta = 0.10$ determine
(a) the steady state optimum (X^*, Y^*), and
(b) the optimal approach paths $\{Y_t^*\}, \{X_t^*\}$.

Compare this to the MSY solution. Why are the two so close?

Answer

(a) $X^* = 47.41$, $Y^* = 449.87$

(b) MRAP is optimal. Since $X_0 > X^*$ and $Y_t = Y_{\max}$ reduces X_{t+1} as rapidly as possible we set $Y_t = 453$ for $t = 0, 1, \ldots, 6$ resulting in

t	1	2	3	4	5	6	7
X_t	97	72.46	62.99	57.53	53.63	50.41	47.41

The MSY solution is $X^* = 50$, $Y^* = 450$. This is close to the discounted case because the intrinsic growth rate $G'(0) - 1 = 99 >> \delta$.

P1.7.8. Use dynamic programming to solve the following problem for $T = 0, 1, 2,$ and 3.

$$\text{maximize} \quad \sum_{t=0}^{T} \rho^t Y_t$$

$$\text{subject to} \quad X_{t+1} = G(X_t - Y_t) \quad t = 0, 1, \ldots, T-1$$

$$0 \le Y_t \le X_t$$

$$\text{given} \quad \rho = 1/(1+\delta), \qquad \delta = 0.1 \qquad X_0 = 1000$$

$$\text{and} \qquad G(S) = \frac{2S}{1 + .001S}$$

If $J_n(X_0) = \max \sum_{t=0}^{n} \rho^t Y_t$ given X_0, show that

$$J_{n+1}(X_0) = X_0 + \max_{0 \le S \le X_0} [\rho G(S) - S]$$

Answer

(a) $T = 0$: $S_0^* = 0$, $Y_0^* = 1{,}000$, $J_0 = 1{,}000$

(b) $T = 1$: $S_0^* = 348.4$, $S_1^* = 0$, $Y_0^* = 651.6$ $Y_1^* = 516.76$, $J_1 = 1{,}121.38$

(c) $T = 2$: $S_0^* = S_1^* = 348.4$, $S_2^* = 0$, $Y_0^* = 651.6$, $Y_1^* = 168.36$, $Y_2^* = 516.76$, $J_2 = 1{,}231.73$.

(d) $T = 3$: $S_0^* = S_1^* = S_2^* = 348.4$, $S_3^* = 0$, $Y_0^* = 651.6$, $Y_1^* = Y_2^* = 168.36$, $Y_3^* = 516.76$, $J_3 = 1{,}332.04$

P1.7.9. What is the value of $J_\infty(X_0)$, $X_0 = 1{,}000$, for the preceding problem?

Answer. $J_\infty = 651.6 + 168.36(\rho + \rho^2 + \cdots) = 2{,}335.20$.