ECONOMICS
RESEARCH
REPORT

GROWTH AND HARVEST WITHOUT CULTIVATION: AN INTRODUCTION TO DYNAMIC OPTIMIZATION

THOMAS JOHNSON

ECONOMIC RESEARCH REPORT NO.48
DEPARTMENT OF ECONOMICS AND BUSINESS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA

ERR 48 JUNE 1985

GROWTH AND HARVEST WITHOUT CULTIVATION: AN INTRODUCTION TO DYNAMIC OPTIMIZATION

THOMAS JOHNSON

Economic Research Report No. 48
Department of Economics and Business
North Carolina State University at Raleigh

June, 1985

ABSTRACT

This report presents the optimal control approach to dynamic optimization. The presentation begins with a simple two-period problem and a level of analysis that should be familiar to anyone who has had an intermediate level course in price theory. The form of this problem is changed slightly to lead smoothly to the development of the Maximum Principle of optimal control for many discrete time periods. This discrete time example is presented in a way that shows clearly the meaning of the Maximum Prinicple in continuous time. An example closely related to the examples given for optimization in discrete time is used to introduce the fundamentals of optimal control in continuous time and the use of phase diagrams for describing important characteristics of the solution.

Appendixes provide a unified treatment of constrained optimization, nonlinear programming, and generalize the statement of the Optimal Control problem from the particular examples presented in the body of the report.

TABLE OF CONTENTS

	Page
I. Introduction	4
II. A Proverbial Example: The Curious Crop	7
A. A Very Simple Problem	7
B. More: More Periods, More Variables, More Constraints	16
C. Continuous Growth and Harvest: Another Curious Crop	17
III. Use With Caution	29
Footnote	31
References	32
Appendix A. Principles of Constrained Optimization	A-1
A. A Review of Optimization	A-1
B. Maximization Subject to an Equality Constraint	A-4
C. Limits on Activities (Variables x_1, x_2)	A-10
D. Inequality Constraints on Resources	A-19
E. An Example of Maximization Subject to Inequality Constraints	A-23
F. More Variables and Constraints	A-28
Footnote to Appendix A	A-30
Appendix B. Representing the Problem	B-1
Footnote to Appendix B	B-6
Appendix C. Discrete Time Maximum Principle (Fixed End Time)	C-1
Appendix D. The Maximum Principle with Continuous Time (Fixed End Time)	D-1

GROWTH AND HARVEST WITHOUT CULTIVATION: AN INTRODUCTION TO DYNAMIC OPTIMIZATION

There are more things in heaven and earth,
Horatio, than are dremt of in your philosophy.

Shakespeare, Hamlet, I.v.166

I. Introduction

To achieve the best results in a world in which events occur through time one must dream of things not yet possible. And, of course, in such speculations there can be no certaintly. But the first step to mastering the dynamic decision process is to solve a problem for which timing is important but for which risk and uncertainty may be ignored.

The purpose of this report is to develop understanding of a technique of analysis that has become very important in agricultural and resource management. Many recent theses and articles in agricultural economics and planned research reports are based on this technique of dynamic analysis. But, at the beginning of 1985 available articles and texts explaining this technique require a knowledge of mathematics well beyond that required for more traditional analyses and beyond the mathematical training of many who will find the technique useful. However, a practical understanding can be obtained using only elementary calculus and algebra. This report will build upon this minimum level of mathematics training. Furthermore, the key concepts can be explained to anyone who can read a contour map and has ever gone a bit out of his way to take a major highway rather than to go the shortest distance on back roads.

To reduce the complexity to a useful minimum, I first consider a crop that is curious indeed: one that is sown by nature and grows without cultivation but that is costly to harvest and changes value over time making the timing of harvest important. One advantage of beginning the study with this curious crop is that it avoids the worry of whether the formulas give a good picture of some real commodity we have in mind. Think of this as drawing an exact diagram of a mythical plant such as Jack's wonderful beanstalk.

The primary advantage of the analysis of this curious crop is that it gives a good context for a statement of the Maximum Principle. Lest you be discouraged by such an awesome thing as "THE" Maximum Principle, consider this homey, but accurate, version:

> You can do as well as possible over the long run just by doing your best at each moment, IF you take proper account of how your actions change your resources.

With this you have "only" to determine what "proper account" to take of changes in your resources and you are back to solving a problem with which you are familiar.

A further advantage of the curious crop is that it gives a sketchy but recognizable picture of a real commodity. Optimal control of the harvest of bay scallops in the sounds of North Carolina is a real problem with only a few complications added to that of the harvest of the curious crop. A detailed account of the application of this kind of model to the North Carolina Bay Scallop Fishery will be contained in a future research report. Other problems that might be well represented by a model very similar to that of the curious crop include the optimal grazing of a forage crop, the optimal market strategy for a potential plant variety, and the optimum control of soil erosion. Many readers may find that the best use of the curious crop example is to make accessible the many more complex analyses continually appearing in the literature.

This report closes with a discussion of the way that such model results should be used: with caution and common sense. The main thing to remember is that people are doing what they do because to the best of their knowledge that is what is best for them. It is very likely that at one time there were good reasons for them to begin what have become traditional practices. When there are major differences between the results presented by a model and established practice, it is very important to find out why those differences occur. Perhaps conditions have changed since the practice was established. But, the difference might reveal a major failure of the model to represent the real problem.

II. A PROVERBIAL EXAMPLE: THE CURIOUS CROP

The basics of optimal control using the maximum principle with both discrete time intervals and continuous time flow are given in the appendices. Insight into many of the most important principles can be gained with a very simple, two-period example. This example doesn't model any real commodity very well but seems close enough to reality to help make the fundamentals of optimal control more clear and appealing. A slightly more realistic parable in continuous time adds the details to show the connection between the discrete time problem and continuous time problems.

A. A Very Simple Problem

Once there was a man who owned a tract of forest land through which a highway was scheduled to be cut in two years. In the path of the road he discovered a stand of commercially valuable herb that could be harvested only one time each year and could not be successfully transplanted or artificially cultivated. When discovered, there were K_0 kilograms of the herb just ready for harvest. The decision to be made was how much to harvest then and how much to leave to produce next year's crop.

The owner called his county agent who called the state extension specialist and learned that 1 kilogram left unharvested this year would grow to more than 1 kilogram next year. With a growth rate of γ per year, 1 kilogram now grows to $(1 + \gamma)$ kilograms next year. The specialist also said that the net price per kilogram, after harvesting costs, would depend on the inventory remaining because as the plants become more rare they become more costly to find and harvest. To express this dependence he wrote P(K) for the price. Therefore, if he harvested u_1 kilograms in year 1 and u_2 kilograms in year 2, the present value (at the beginning of the project) of net returns in years 1 and 2 would be

$R_1 = P(K_1)u_1$ for year 1 and $R_2 = \frac{P(K_2)u_2}{(1+r)}$ for year 2.

The term 1/(1+r) gives the present value at year 1 when the interest rate is r per year. The objective was to maximize the sum of the present values $R_1 + R_2$. The owner then discovered an additional constraint. The government required that a fraction f of the original stock be left in what was to be the undisturbed part of the roadside. This meant that in year 3 he had to have

$$K_3 \geq fK_0 \text{ (that is, } K_3 - fK_0 \geq 0). \quad \text{(c)}$$

Since the plan for harvesting began in year 1 with the known initial stock, he had $K_1 = K_0$ to begin with. Since each kilogram that was not harvested in one year would grow to $1 + \gamma$ kilograms the next year, he also had the constraints

$$K_2 = (1+\gamma)(\text{amount left after year 1 harvest}) = (1+\gamma)(K_1 - u_1) \quad \text{(a)}$$

$$K_3 = (1+\gamma)(\text{amount left after year 2 harvest}) = (1+\gamma)(K_2 - u_2). \quad \text{(b)}$$

These growth constraints are easier to generalize if they are written in the form of the change in K between years:

$$K_2 - K_1 = g_1(K_1, u_1) \text{ where } g_1(K_1, u_1) = \gamma K_1 - (1+\gamma)u_1$$

$$K_3 - K_2 = g_2(K_2, u_2) \text{ where } g_2(K_2, u_2) = \gamma K_2 - (1+\gamma)u_2.$$

Writing the growth constraints this way shows that the change from one year to the next equals the amount the original stock would have grown (e.g., γK_1) minus the sum of the amount harvested and the amount this harvest would have grown (e.g., $u_1 + \gamma u_1 = (1+\gamma)u_1$). Since the herb cannot be obtained from other sources, the quantities harvested cannot be negative, that is, $u_1 \geq 0$, $u_2 \geq 0$.

The owner's whole problem is now laid out. He wants to solve the following problem:

$$\text{Maximize } R = R_1(K_1, u_1) + R_2(K_2, u_2) \quad (1)$$

with respect to u_1, u_2

subject to

$$K_2 - K_1 = g_1(K_1, u_1) \qquad K_1 = K_o \tag{1a}$$

$$K_3 - K_2 = g_2(K_2, u_2) \tag{1b}$$

$$K_3 - fK_o \geq 0 \tag{1c}$$

$$u_1 \geq 0 \quad u_2 \geq 0. \tag{1d}$$

The quantity sold, u, is the CONTROL VARIABLE the decision maker can manipulate directly in each year. The inventory K is the STATE VARIABLE in each year. The state variable changes as a result of previous values of the control variables and the passage of time but cannot be manipulated directly by the decision maker. For example, K_2 depends on K_1 and u_1 but does not depend on u_2. The TERMINAL CONDITION is an equation that imposes a requirement on the value of the state variable in the year following the last year in the time horizon. Here the terminal condition is $K_3 \geq fK_o$.

Since there are a finite number of years with a finite number of state variables and control variables each year, this problem can be solved in the usual way with Lagrange Multipliers. See Appendix A for a review of optimization. Section C of Appendix A gives the implications of having inequality constraints on activities such as "no negative sales, $u_1 \geq 0$, $u_2 \geq 0$" in this problem. Section D of Appendix A gives the implications of having inequality constraints on resources such as "have at least a fraction f of the original stock left, $K_3 - fK_o \geq 0$."

Since there are only two periods for growth and harvest, this problem may be shown in a two-dimensional graph of the familiar type. To get the problem into a convenient form to diagram, first consider constraint (c), the requirement for the state value in period 3, $K_3 \geq fK_o$. From constraint (b) we know that the amount K_3 is just what was left from the harvest in period 2 (that is, $K_2 - u_2$) after it has grown for one period:

$$K_3 = (1+\gamma)[K_2 - u_2].$$

Also, from constraint (a), the amount K_2 is just what was left from the harvest in period 1 after it has grown for one period (remember that $K_1 = K_0$), so

$$K_2 = (1+\gamma)(K_0 - u_1).$$

Substituting this formula for K_2 into the formula for K_3 gives

$$K_3 = (1+\gamma)[(1+\gamma)(K_0 - u_1) - u_2].$$

Collecting terms in K_0, u_1 and u_2 now gives the constraint (c) for K_3 in the following form:

$$K_3 = (1+\gamma)^2 K_0 - (1+\gamma)^2 u_1 - (1+\gamma)u_2 \geq fK_0,$$

which after collecting terms in K_0 , can be written as

$$(1+\gamma)^2 u_1 + (1+\gamma)u_2 \leq [(1+\gamma)^2 - f]K_0. \quad (2)$$

This constraint is shown in Figure II.1. This figure is labelled to show that if all allowed harvesting is done in period 1, the amount u_1 will be $K_0(1-f/(1+\gamma)^2)$. This will leave $fK_0/(1+\gamma)^2$, which will grow to fK_0 by period 3. If there is no harvest in period 1, the maximum available for harvest will grow to the amount $K_0(1+\gamma)(1-f/(1+\gamma)^2)$, shown as the intercept of the constraint on the u_2 axis.

With the constraints transformed in this way, the problem could be written as maximizing R, the net present value of harvests, subject to the single constraint given by equation 2. However, this form as a single constraint does not easily show how to generalize to more than two periods. What I will do is show the solution with the three constraints in equation 1 and then show how this gives the answer to the more familiar form with the single constraint. The Lagrangian for the problem stated in equation 1 is

$$\mathcal{L} = R_1 + R_2 + \lambda_1(g_1 - (K_2 - K_1)) + \lambda_2(g_2 - (K_3 - K_2))$$
$$+ \sigma(K_3 - fK_0), \quad (3)$$

with $K_1 = K_0$.

This is an especially simple case of the discrete time optimal control problem given in Appendix C. In optimal control problems the λ's are called <u>Adjoint</u> or <u>Costate</u> variables and the Lagrange multiplier, σ, for the terminal constraint is written with a different symbol to show that this is a different kind of constraint.

There are two kinds of "activities" in this problem. One kind is K, the amount of the plant available, and the other is u, the amount of plant harvested. The necessary conditions for optimality require that the partial derivative of the Lagrangian function, $\mathcal{L}$, with respect to each activity equal zero. Also, the partial derivatives with respect to each of the λ's must equal zero since they are associated with equality constraints. In the control literature each of these kinds of conditions has a name that gives some insights into its meaning. These names help one remember all of the conditions and help make the connection between this little discrete time problem and continuous time optimal control. These necessary conditions are as follows:

First, take the partial derivative with respect to the state variable K.

Adjoint Condition

$$\frac{\partial \mathcal{L}}{\partial K_2} = \frac{\partial R_2}{\partial K_2} + \lambda_2 \frac{\partial g_2}{\partial K_2} + \lambda_2 - \lambda_1 = 0$$

or (4)

$$\lambda_2 - \lambda_1 = - \frac{\partial}{\partial K_2}(R_2 + \lambda_2 g_2) .$$

This is called the adjoint equation because it gives the rate of change in the adjoint (costate) variable λ. Since $K_1 = K_o$ is a given constant, K_1 is not a variable and there is no derivative with respect to K_1.

Transversality Condition

$$\frac{\partial \mathcal{L}}{\partial K_3} = - \lambda_2 + \sigma \frac{\partial}{\partial K_3}(K_3 - fK_o) = 0$$

or

$$\lambda_2 = \sigma. \quad \text{Also, } \sigma \geq 0 \text{ and } \sigma(K_3 - fK_o) = 0 . \qquad (5)$$

The additional conditions $\sigma \geq 0$ and $\sigma(K_3 - fK_o) = 0$ come from the Kuhn-Tucker condition (See Appendix A, Section D, and Appendix C) Notice that the constraint $K_3 - fK_0 \geq 0$ is treated as a resource constraint to produce the shadow price σ. This is called the transversality condition because it gives the condition that must be satisfied as the project transverses the planning horizon.

The partial derivatives of the Lagrangian with respect to the multipliers, λ's, reproduce the constraints as the

Equations of Motion

$$\frac{\partial L}{\partial \lambda_1} = g_1 - (K_2 - K_1) = 0$$

$$\frac{\partial L}{\partial \lambda_2} = g_2 - (K_3 - K_2) = 0. \tag{6}$$

These are called the equations of motion because they describe how the resource grows (moves) over time.

Finally take the partial derivatives with respect to the control variables, the quantities harvested u_1 and u_2.

Hamiltonian Condition

$$\frac{\partial L}{\partial u_1} = \frac{\partial R_1}{\partial u_1} + \lambda_1 \partial g_1/\partial u_1 \leq 0 \qquad u_1 \frac{\partial L}{\partial u_1} = 0 \qquad u_1 \geq 0$$

$$\frac{\partial L}{\partial u_2} = \frac{\partial R_2}{\partial u_2} + \lambda_2 \partial g_2/\partial u_2 \leq 0 \qquad u_2 \frac{\partial L}{\partial u_2} = 0 \qquad u_2 \geq 0. \tag{7}$$

The reason for calling these the Hamiltonian Condition will be seen soon when the problem is given in a different form using the Hamiltonian function instead of the Lagrangian. The inequalities enter these conditions because the harvest rates cannot be less than zero. The equation $u_1(\partial L/\partial u_1) = 0$ says either $u_1 = 0$ or $\partial L/\partial u_1 = 0$. If the optimal harvest rate is positive, then the usual equality constraint $\partial L/\partial u_1 = 0$ holds.

The form of these necessary conditions can be simplified and the MAXIMUM PRINCIPLE can be obtained with the following definition. For this example the HAMILTONIAN FUNCTION at each time period i is defined to be

$$H_i(k_i, u_i, \lambda_i) = R_i + \lambda_i g_i \ .$$

THE HAMILTONIAN FUNCTION IN PERIOD i IS THE NET RECEIPTS FROM SALES ("PROFIT" R_i) PLUS THE VALUE OF THE ADDITIONS TO INVENTORY VALUED AT THE SHADOW PRICE λ_i.

The partial derivative of H_i with respect to u_i is

$$\frac{\partial H_i}{\partial u_i} = \frac{\partial R_i}{\partial u_i} + \lambda_i \frac{\partial g_i}{\partial u_i} = \frac{\partial L}{\partial u_i} \ .$$

This means that the conditions of equation (7) are necessary to maximize the Hamiltonian at each period. Notice also how H can be used to write equations (4) and (6) more compactly.

The problem of maximizing over <u>all periods</u>, given in equation 1, will be solved if the following problem is solved <u>for each period i</u>.

Maximize $H_i(K_i, u_i, \lambda_i)$

w.r.t. u_i

together with

Adjoint Condition (from 4)

$$-\frac{\partial H_2}{\partial K_2} = (\lambda_2 - \lambda_1)$$

Transversality Condition (from 5)

$$\lambda_2 = \sigma\frac{\partial}{\partial K_3} (K_3 - fK_o)$$

Equations of Motion

$$\frac{\partial H_1}{\partial \lambda_1} = K_2 - K_1$$

$$\frac{\partial H_2}{\partial \lambda_2} = K_3 - K_2$$

Initial Condition

$$K_1 = K_o$$

THE MAXIMUM PRINCIPLE SAYS THAT THE RESOURCE MANAGER CAN GET OPTIMAL RESULTS OVER THE WHOLE PLANNING HORIZON IF THE BEST IS DONE AT EACH MOMENT. THE CATCH IS THAT TO KNOW WHAT IS BEST TO DO AT EACH MOMENT, THE MANAGER MUST KNOW THE SHADOW PRICE, λ_i, FOR A CHANGE IN THE STATE VARIABLE, K.

For a more rigorous and general statement of the Maximum Principle see Kamien and Schwartz (pp. 201-203), Intriligator (p. 351) and Clark (p. 91, p. 251). The shadow prices are connected over the entire horizon by the Adjoint Condition and the Transversality Condition.

The Transversality Condition (5) says that the marginal value of the resource saved to meet the government's requirement for the remaining stock must equal the marginal value of the resource in the last period of harvest. This also says that if K_3, the amount of the resource in year 3, is more than the government's required value, fK_0, then $\lambda_2 = \sigma = 0$, which means there is no marginal payoff to leaving this much of the resource in the last period. Harvest should continue until either the net return goes to zero or the constraint on K_3 becomes binding.

At first sight, the negative sign in the Adjoint Condition, $-\partial H_2/\partial K_2 = \lambda_2 - \lambda_1$, is puzzling because it says that the greater the current marginal contribution of the stock K, the more rapidly the shadow price λ must decrease. The Adjoint Condition is written in the puzzling form here because this is the form in which the Adjoint Condition appears in the continuous time problem and the puzzle is more easily solved with the discrete time problem. This puzzle is solved by noticing that the shadow price must be moving toward the end value given by the Transversality Condition, $\lambda_2 = \sigma$. The shadow price decreases more rapidly because it starts from a higher value when the marginal contribution of K_2 is higher. This effect is clear when the Adjoint Condition is written as $\lambda_1 = \lambda_2 + \partial H_2/\partial K_2$, from which it is clear that a greater marginal contribution of the stock must increase the earlier shadow price.

Now in the Adjoint Condition (equation 4), substitute $\lambda_2 = \sigma$ from equation (5) and $\partial g_2/\partial K_2 = \gamma$ from the definition of g_2 to get

$$\lambda_1 = \sigma_2 + \gamma\sigma_2 + \partial R_2/\partial K_2$$

or

$$\lambda_1 = (1+\gamma)\sigma + \partial R_2/\partial K_2. \qquad (9)$$

To find λ_1, the marginal value of the resource in period 1, I had to go to the end, find out what the marginal value would be for the terminal constraint, and work back to the beginning. But when I know the λ's, the Maximum Principle tells me how to get the best results for all time by just doing my best one period at a time.

As promised, I now return to the familiar form of the problem with one constraint to reassure the reader that the two different forms give the same solution. First the problem in the form shown in Figure 1 is given with the constraint in the form in equation 2, by

$$\text{Max } R_1(K_1, u_1) + R_2(K_2, u_2)$$
$$\text{w.r.t. } u_1 \text{ and } u_2$$

subject to

$$(1+\gamma)^2 u_1 + (1+\gamma)u_2 \leq [(1+\gamma)^2 - f)]K_o.$$

Remember that $K_1 = K_o$, $K_2 = (1+\gamma)(K_o - u_1)$. The Lagrangian for this problem is

$$L_1 = R_1 + R_2 + \mu\{[(1+\gamma)^2 - f)]K_o - (1+\gamma)^2 u_1 - (1+\gamma)u_2)\}$$

and the necessary conditions, when both u_1 and u_2 are greater than zero and the constraint is an equality, are

$$\frac{\partial R_1}{\partial u_1} + \frac{\partial R_2}{\partial K_2}\frac{dK_2}{du_1} - (1+\gamma)^2\mu = 0$$

$$\frac{\partial R_2}{\partial u_2} - (1+\gamma)\mu = 0,$$

with $dK_2/du_1 = -(1+\gamma)$. Equating these two ways of writing zero gives

$$\frac{\partial R_1}{\partial u_1} - (1+\gamma)\frac{\partial R_2}{\partial K_2} - (1+\gamma)^2\mu = \frac{\partial R_2}{\partial u_2} - (1+\gamma)\mu$$

or, since $(1+\gamma)^2 - (1+\gamma) = \gamma(1+\gamma)$,

$$\frac{\partial R_1}{\partial u_1} = \frac{\partial R_2}{\partial u_2} + (1+\gamma)\frac{\partial R_2}{\partial K_2} + \gamma(1+\gamma)\mu. \qquad (10)$$

Now from the Hamiltonian equation (7), the relation between σ, λ_2 and λ_1 from equations (5) and (9), and the fact that $\partial g_1/\partial u_1 = \partial g_2/\partial u_2 = -(1+\gamma)$, when the optimal u_1 and u_2 are positive,

$$\frac{\partial R_1}{\partial u_1} = [(1+\gamma)\sigma + \partial R_2/\partial K_2](1+\gamma) = \frac{\partial R_2}{\partial u_2} - \sigma(1+\gamma)$$

or

$$\frac{\partial R_1}{\partial u_1} = \frac{\partial R_2}{\partial u_2} + (1+\gamma)\frac{\partial R_2}{\partial K_2} + \gamma(1+\gamma)\sigma. \qquad (11)$$

Comparison of equations (10) and (11) shows that the two methods give identical results with $\mu = \sigma$ when the single constraint is in the form in equation (2).

The more familiar form of the resource allocation problem, with the constraint in the form of equation (2) and shown in Figure 1, can also be generalized to continuous time, more constraints, etc. (see Kamien and Schwartz, pp. 212-214 for an example). But for problems with many time periods or continuous time, the generalization of the form that has given the Maximum Principle is usually the more convenient form.

B. MORE: More Periods, More Variables, More Constraints

The very simple problem with only one resource, one control variable, and two time periods is easily generalized using the knowledge of optimizing techniques from Appendix A. Appendices B and C give the form of the problem and equations for applying the Maximum Principles to problems with T discrete time intervals, n control variables in each period, m state variables, r constraints on the control variables, and s terminal constraints. It sounds terribly complicated, and real research problems can require models that increase any or all of the numbers n, m, r and s. But it is amazing how well the simple two-period problem above prepares one to tackle the larger problems.

In the simple problem with two periods and one control each period, the Maximum Principle changed it into two problems each with one control variable. The reader probably has already concluded that the Maximum Principle is not worth the trouble when you have only a two-period problem. In a problem with T discrete time periods and n controls each period, the usual formulation with a Lagrangian gives one problem with nT controls to be determined. The Maximum Principle approach with the Hamiltonian for each period changes this into T problems, each with n controls. This still may not be such a bargain until the number of periods T becomes very large. For problems with continuous time, however, the "number of periods" becomes infinite so instead of trying to solve one problem in infinite dimensions, the Maximum Principle changes the problem to an infinite number of problems, each with a finite number of dimensions. That _is_ a bargain because an ordinary function of time can present the solution for all values of time. As soon as students of elementary algebra find a solution for y as a function of x they know how to solve an infinite number of problems of the form "for this value of x what is the value of y?" It is in this sense that an infinite number of problems are solved in the Maximum Principle.

C. Continuous Growth and Harvest: Another Curious Crop

Meanwhile, back in the forest, the highway has been completed and the herb harvest has ended. However, the landowner has discovered a unique variety of ornamental plant on which he has obtained a patent. The patent is perfectly effective for a total of T years, after which there will be no profit in raising this plant. Fortunately the owner has all the greenhouse space he can use, a ready market, and an initial parent stock of K_0 kilograms. This plant grows year-round in greenhouses with a rate of increase g(K), depending upon the mass K of parent plants under cultivation. The growth function $g(K)$ is a continuous, differentiable function that changes only with K and not directly with time. The growth function also has the properties $g(0) = 0$, $g'(0) > 0$, and $g''(K) < 0$ where g' denotes the first derivative and g'' denotes the second derivative. These mathematical conditions mean that there is no growth without some parent stock, $g(0) = 0$; that for small quantities of the parent stock the growth rate increases with increases in the parent stock, $g' > 0$ for $K = 0$; and the increase in the rate of growth decreases with increasing parent stock, $g''(K) < 0$ for all values of K.

I like to call this plant Crusonia Curiosa because of its similarity to the equally unreal Crusonia plant used as an example by Knight (1944).

The harvest is in the form of cuttings that are sold by weight. This is the control variable that is a continuous function of time written as $u(t)$. Since the plant is harvested at the moment of sale, there is no distinction between the rate of harvesting and the rate of selling. For this problem I assume it is always optimal to have positive harvest. The effects of constraints on u are shown in Appendix D. Because of the patent monopoly on the sales of this plant, revenue decreases as the rate of selling increases, $(R')' = R'' \leq 0$, even though marginal revenue is positive, $R' > 0$.

The owner's objective is to maximize the present value of the stream of net revenue over the life of the patent, and the interest rate for discounting the future is r. As we go from discrete time intervals to continuous passage of time, the discount factor $1/(1+r)^n$ becomes e^{-rt}. As usual, this rate may be interpreted as the opportunity cost of credit with continuous compounding. There is no "scrap value" to inventory remaining after the patent ends. Using the notation $\vec{u}(t)$ to mean the entire path of harvest over time t between 0 and T, the manager's problem is given by the following mathematical statement:

$$\text{Maximize} \qquad \int_{t=0}^{T} R(u(t))e^{-rt}dt$$

with respect to $\vec{u}(t)$

subject to

$$\dot{K} = g(K) - u(t)$$

$$K(0) = K_0. \quad K(T) \geq 0. \tag{12}$$

The net rate of increase, $\dot{K} = dK/dt$, in the stock parent plant (inventory) equals the rate of growth of the plant, $g(K)$, minus the rate of harvest, which equals the rate of sales.

The Hamiltonian for this problem is

$$H(t) = R(u(t))e^{-rt} + \lambda(t)(g(K) - u). \tag{13}$$

By the Maximum Principle (see Appendix D for the transition to the continuous time case and Appendix C for an explanation of the Hamiltonian), with the conditions on the derivatives and second derivatives of R and g, the necessary and sufficient conditions to solve the problem of equation 12 are as follows:

At every time t from 0 to T

Maximize $H(t)$

with respect to $u(t)$

together with (14)

$$\dot{\lambda} = -\frac{\partial H}{\partial K} \qquad \lambda(T) \geq 0 \qquad \lambda(T)\,K(T) = 0$$

$$\dot{K} = \frac{\partial H}{\partial \lambda} \qquad K(0) = K_0 \qquad K(T) \geq 0\,.$$

Pause a moment to reflect on the meaning of each of the elements in equations (13) and (14). Remember that the adjoint variable, also called the costate variable, $\lambda(t)$, gives the marginal value of the state variable K at time t. Therefore, the quantity $\lambda(g - u)$ is the marginal value of the rate of increase of the parent plant. Of course, if the rate of harvest, $u(t)$, is greater than the rate of growth, $g(K(t))$, then $\lambda(g - u)$ is negative or zero since λ must be greater than or equal to zero. The quantity Re^{-rt} is the direct contribution at time t to the objective which is to maximize the integral of Re^{-rt} over the planning horizon. So the Hamiltonian $H(t)$ is the sum of the direct and indirect contributions of the harvest at time t.

Also notice the similarities and differences in the Adjoint Equation, $\dot{\lambda} = -\partial H/\partial K$, and the Equation of Motion, $\dot{K} = \partial H/\partial \lambda$. The negative sign in the Adjoint Equation says that the greater the marginal contribution of the stock of the parent plant to the value of the Hamiltonian objective function, the more rapidly the shadow price must be decreasing with time. Recall how this puzzle was solved in the discrete time period problem that gave equation (4): a given end value and more rapid decreases in the shadow price imply higher earlier values. The boundary condition on the shadow price of the parent stock is given at the end of the planning horizon. Having this end value condition on the costate variable makes the equations more difficult to solve than they would be if the initial value were known, as it is for the stock of parent plant, K_0. But remember that the shadow price contains the information about the future value of the stock. The state variable K tells us about the past, the costate variable λ tells us about the future.

Assume that the optimal harvest rate, u, will be greater than zero but not large enough to exhaust the stock instantaneously. Then the necessary condition $\partial H/\partial u = 0$ together with the equations for $\dot{\lambda}$ and $\dot{K}$ from equation 14 gives the following necessary conditions to solve the original management problem given in equation 12.

<u>At each t</u>

(a) $\frac{dR}{du}e^{-rt} - \lambda = 0$

(b) $\dot{\lambda} = -\lambda\frac{dg}{dK}$ $\qquad \lambda(T) \geq 0 \qquad \lambda(T)\,K(T) = 0$

(c) $\dot{K} = g - u$ $\qquad K(0) = K_0 \qquad K(T) \geq 0.$ (15)

If the functional forms for R(u) and g(K) were simple enough, the conditions of equation (15) could be solved for u, K, and λ as functions of t. Or numerical integration with a computer could be used to get as close approximations as required for any reasonable functional forms of R and g. But, through use of the technique of "phase diagrams" some important characteristics of solutions can be seen without getting the final solutions of the equations. Building up these phase diagrams is important for the study of economics or resource management because they help use our knowledge of static problems to help understand the dynamic. Also, understanding these phase diagrams gives insights into the actual solution of the management problem, whether this solution be a numerical approximation obtained with a computer or a less formal approximation based on experience and judgment without extensive calculations.

For convenience in notation, let $R' = dR/du$, $R'' = d^2R/du^2$, $g' = dg/dK$, $\dot{\lambda} = d\lambda/dt$, $\dot{u} = du/dK$. From equation (15a)

$$\lambda = R'e^{-rt} \tag{16}$$

and differentiating λ from this expression with respect to t gives (remember to use the chain rule since the control u is also a function of time)

$$\dot{\lambda} = -rR'e^{-rt} + R''\dot{u}e^{-rt}.$$

Now factor $R'e^{-rt}$ out of the right-hand side,

$$\dot{\lambda} = R'e^{-rt}(-r + \frac{R''}{R'}\dot{u}).$$

Recognize the value for λ from equation (16), so

$$\dot{\lambda} = \lambda(-r + \frac{R''}{R'}\dot{u}) . \tag{17}$$

Equate the expressions for $\dot{\lambda}$ from equations (15b) and (17) and divide by $-\lambda$ to get the following equation that does not involve λ:

$$g'(K) = r - \frac{R''}{R'}\dot{u} ,$$

which can be solved for $\dot{u}$ to give

$$\dot{u} = \frac{R'}{R''}(r - g'(K)) , \tag{18}$$

assuming that R'' is not equal to zero. If R is a linear function of u, then R'' is zero and the optimal harvest rate is either zero or its maximum value (so-called bang-bang control). It is clear from equation (18) that the optimal harvest rate is constant, $\dot{u} = 0$, if and only if

$$g'(K) = r , \tag{19}$$

which implicitly gives a critical value of K, call it K_c, for which the harvest rate is constant. This value is diagrammed as the vertical line labeled $\dot{u} = 0$ in Figure II.2.

Reasoning through from equation (18), it becomes clear that the optimal harvest rate is decreasing through time, $\dot{u} < 0$, if the stock is greater than the critical value, i.e., if $K > K_c$. The reasoning goes this way. Since marginal revenue is positive but decreasing with sales, $R' > 0$ and $R'' < 0$, the sign of the optimal rate of change in harvest (and sales), $\dot{u}$, is the opposite of the sign of $r - g'$, i.e., the sign of $\dot{u}$ is the same as the sign of $g' - r$. It has been assumed that the second derivative of the growth rate is negative, $g'' < 0$, which means that g' is a decreasing function of K. Therefore, when K is greater than K_c (where $g' = r$), it must mean that the optimal value of g' is less than r, making $\dot{u} < 0$. This fact is indicated by the arrows pointing down in Figure II.2. Similarly, if the stock is less than the critical value, $K < K_c$, then the optimum requires $\dot{u} > 0$. This is indicated by the arrows pointing up in Figure II.2. Together these three relationships between the rate of change in the optimal harvest rate u and the quantity of the parent plant K give the first part of a PHASE DIAGRAM.

The relationships between the rate of change in the stock K and the optimal harvest rate u give the second part of a PHASE DIAGRAM. To find the relationship for this problem, return to the equation of motion that shows the net growth rate

$$\dot{K} = g(K) - u.$$

This says that the inventory of parent plant is constant over time if and only if the harvest rate u equals the growth rate g,

$$u = g(K). \qquad (20)$$

If the harvest rate u is greater than g, then the inventory must be decreasing, $\dot{K} < 0$; and if the harvest rate is less than g, then the inventory must be increasing, $\dot{K} > 0$. By assumption, there is no growth without some parent stock, $g(0) = 0$, and the g function is concave, $g'' < 0$. Therefore, the curve of the harvest rate u as a function of K that would keep the inventory constant, $u = g(K)$, has the form shown

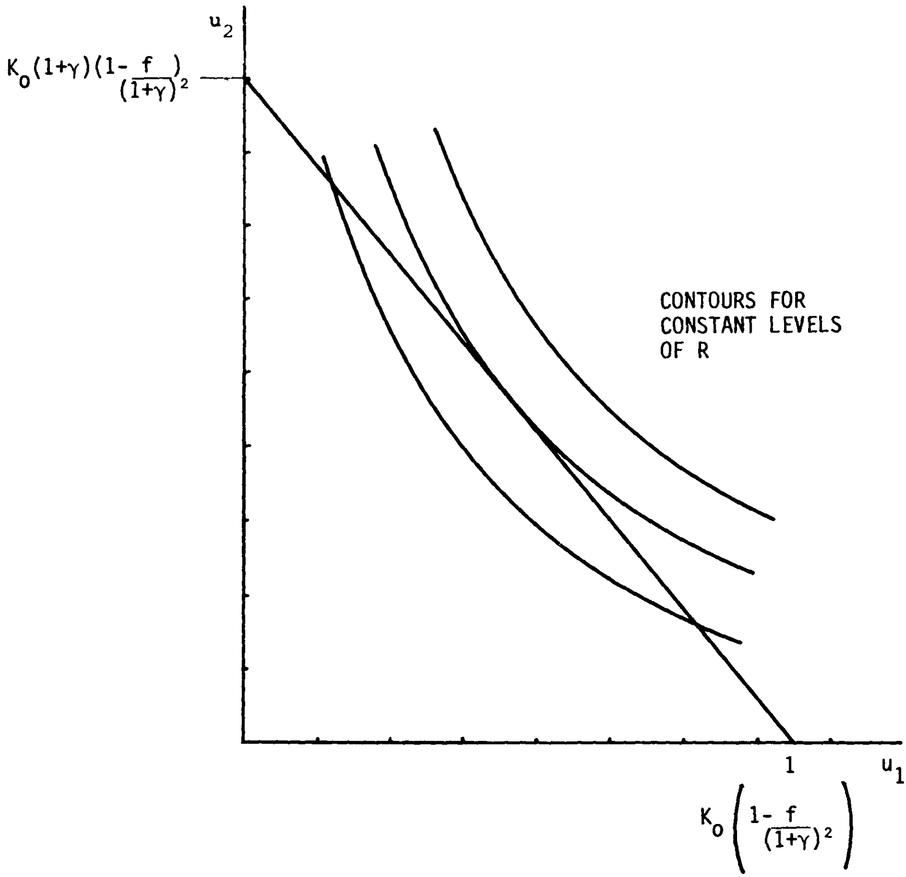


Figure II.1 Terminal constraints as a constraint on harvest

Figure II.2. Direction of change in optimal harvest u

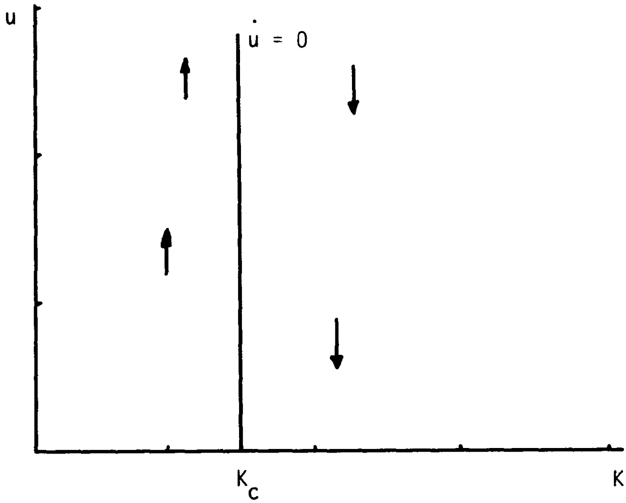
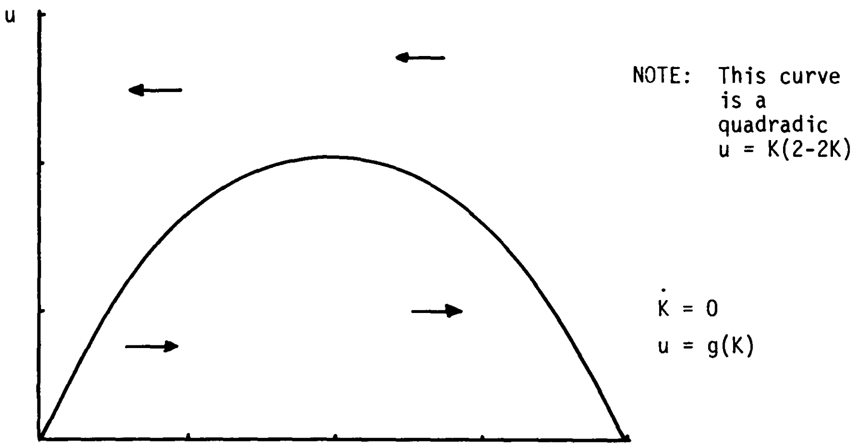


Figure II.3. Direction of net growth in parent stock K



NOTE: When combined these two figures give Figure II.4.

in Figure II.3.[1] The arrows show the direction of change of K with time in the two regions separated by the curve defined by $u = g(K)$. This is the second part of a PHASE DIAGRAM.

When combined, the information in Figures II.2. and II.3. give the PHASE DIAGRAM for the problem of continuous growth and harvest given by equation (12). This is shown in Figure II.4. In each of the four regions (phases) of Figure II.4., the harvest rate u and the inventory level K change over time as indicated by the arrows. For example, in the lower left-hand region, both u and K must be increasing to satisfy the necessary conditions for an optimal plan. Note also that at the value K_c, the slope of the curve for $\dot{K} = 0$ must equal the interest rate, r, $g'(K_c) = r$. This has important implications that will be discussed after looking at some representative paths of u and K through time.

Figure II.5. shows some representative time paths of u and K implied by the phase diagram for two different initial values of K and three different starting values of u for each initial value of K. Notice that when a path crosses one of the boundary curves ($\dot{u} = 0$ or $\dot{K} = 0$), the path must be perpendicular or horizontal at the boundary curve. This is because paths represent motion through time and the boundaries give points where the indicated variable is stationary in time, while the other variable continues to change. Therefore, on the boundary $\dot{K} = 0$, K is not changing with u; on the boundary $\dot{u} = 0$, u is not changing with K.

Paths A, B, and C illustrate available paths satisfying all of the necessary conditions from equation (15) EXCEPT the requirement that $\lambda(T) \geq 0$, $\lambda(T)\, K(T) = 0$. The optimal path among those available for the given K_0 is one that satisfies these conditions on $\lambda(T)$. Equation (15a), $R'e^{-rt} = \lambda$, shows the connection between the shadow price λ and the optimal harvest rate. If the initial λ is chosen too small, this requires R' to be too small, which means that u is too large. Such a choice will exhaust the parent population too soon and will cause λ to become negative before the end time T (λ must not be negative). Similarly, choosing an initial value of λ that is too large causes the value of the harvest rate to be too small and both

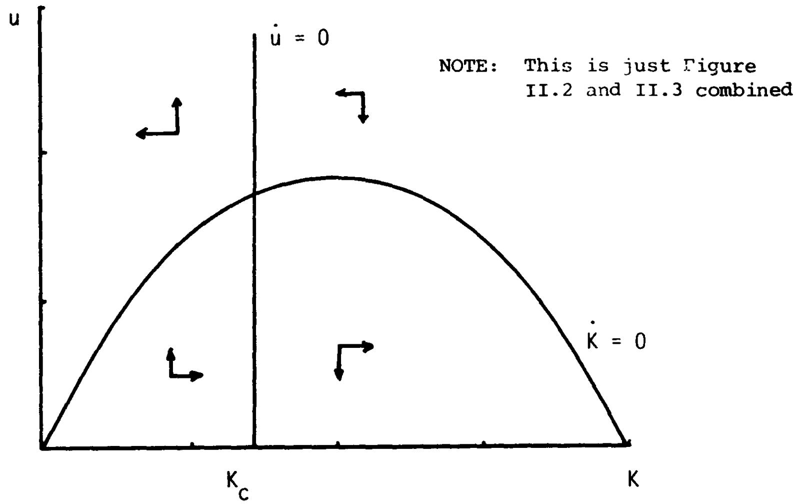


Figure II.4 Optimal Harvest and net growth phases

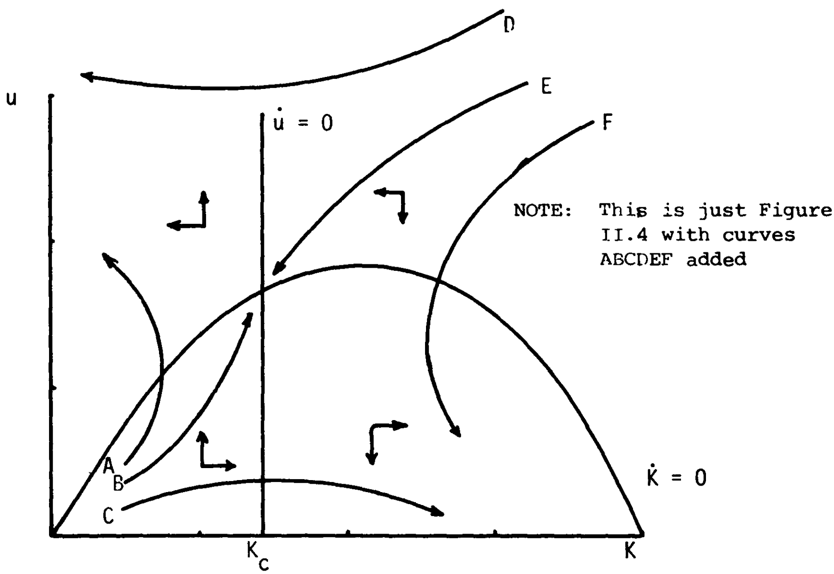


Figure II.5 Representative time paths of harvest rates and stock levels implied by phase diagram

$\lambda(T) > 0$ and $K(T) > 0$; the manager acts as if the parent stock was worth something at the end of the planning horizon.

Path C clearly is not optimal no matter how long the planning horizon. On path C the harvest rate goes to zero and the inventory of parent stock will reach the highest values allowed by the environment (where the $\dot{K} = 0$ curve crosses the K axis). Clearly, patch C does not maximize the present value of net revenues. On the other hand, as the planning horizon goes to infinity, a path such as B will be optimal. Path B approaches the equilibrium point where $\dot{K} = \dot{u} = 0$ and nothing changes.

The equilibrium point where $\dot{K} = \dot{u} = 0$ is the point of <u>optimum</u> sustained yield. Recall from Figure II.4. that at the point at which $\dot{K} = \dot{u} = 0$, the slope of the $\dot{K} = 0$ curve, $u = g(K)$, must equal the interest rate. Therefore, as the interest rate goes to zero, the equilibrium point goes to the highest point on the $\dot{K} = 0$ curve where $g'(K) = 0$. This highest point on the $\dot{K} = 0$ curve is the point of <u>maximum</u> sustained yield. Some have proposed that maximum sustained yield be the criterion for harvest of resources such as fish. However, this is an extreme policy that is optimal only when the interest rate is zero and the planning horizon is infinite (see Clark, p. 43).

For a finite horizon, $T < \infty$, some path such as A is optimal. There are several things to notice about the optimal path with a finite horizon. With a finite horizon, the optimal harvest rate is always at least as great as it would be for an infinite horizon starting with the same initial stock, K_0. It <u>may</u> be optimal to start with a low enough harvest rate to allow the stock to increase for a while. It will always be optimal to harvest rapidly enough to cause the stock to decrease before the end time T. It may be optimal to harvest rapidly enough to cause the stock to decrease from the start (starting with $K_0 < K_c$ such a path would start above the $\dot{K} = 0$ value of u). Whether it will be optimal to exhaust the stock on or before the end time T cannot be determined without more knowledge about the returns function, R(u). If harvest costs increased with decreasing parent stock, then there might be a level of stock that would not be profitable to harvest. In such a situation, both the shadow price, λ, and the harvest rate, u, might optimally go to zero before the end of the horizon and some parent stock might be left unharvested.

If the initial stock K_0 is greater than the critical value K_c, the available paths are represetned by paths D, E and F. Path F would never be optimal for the same reason as that discussed for Path C. For an infinite planning horizon, Path E would be optimal, whereas for a finite horizon, some path such as D is optimal. For this problem, the optimal path must lead to $\lambda(T) = 0$ and the parent stock inventory K will go to zero exactly at the end time T since marginal revenue, R', is positive for all levels of the parent stock.

III. USE WITH CAUTION

The Maximum Principle is a powerful tool, both in concept and in application. As a model of the behavior of people in a dynamic world, it shows that one seeking optimum results will account for the value resources will have in future use. But the analysis shows clearly that the planning horizon, T, and the discount rate, r, have important influences on the value of resources to the decision maker. The equations of motion, which represent the growth and decay of resources, are now important elements of the technology, as they limit the rate at which the world can be changed. More usual production costs, as well as the benefits sought, enter the objective function.

The modeling assumption that the objective function is additive and separable across time may require that some additional effects be treated as resources. Knowledge, skills, likes and dislikes may be important to treat as "resources" to be built up or torn down. It is in this way that a stock of experiences may be modeled to affect decisions. The difficulty is that many of the important elements may not be directly measurable or can be measured only in situations other than the exact setting of the problem. An important example of such a measurement problem is the fact that the growth response of a plant or animal is best measured in experiments not directly connected with the specific decision problem. On the other hand, human capital developed through schooling and experience cannot be observed directly and, therefore, effects must be inferred from the results of the decision process.

In empirical applications estimating parameters and/or testing hypotheses, there often are difficult statistical problems at or beyond the frontiers of econometric methods. Particularly difficult problems arise when the dynamic model includes corner solutions in which control or state variables reach some limit.

Random effects may enter dynamic models in several important forms. The easiest form to handle is that in which the planning horizon or lifetime T is a random variable with a well behaved distribution. There also are known methods for analysis of problems in which the equations of motion are stochastic processes. See for example, Malliaris and Brock (1982). But there are still many difficulties in estimation and hypothesis testing in models in which the equations of motion are stochastic processes in continuous time.

As stated in the introduction, the main thing to remember is that people are doing what they do because to the best of their knowledge that is what is best for them. When there are major differences between the results of a model and established practice, it is very important to find out the reasons for these differences. Perhaps conditions have changed since the practice was established, but, on the other hand, these differences may reveal a major failure of the model in representing the real problem.

Footnote

[1]The quadratic form of g(K) that meets the requirements $g(0) = 0$, $g'' < 0$ is $g(K) = aK - bK^2$. With no harvest, $u = 0$, this gives the logistic growth function in K. In integrated form the Logistic function obtained from $\dot{K} = aK - bK^2$ is

$$K = (\frac{a}{b})/(1 + \frac{\frac{a}{b} - K_o}{K_o} e^{-at}),$$

where a/b is the maximum value K approaches as t goes to infinity and K_o is the value of K at $t = 0$.

REFERENCES

Baumol, W. J., Economic Theory and Operations Analysis, Fourth Edition. Prentice-Hall, 1977, pp. 140-176.

Cannon, M. D., Cullum, C. D. and Polak, E., Theory of Optimal Control and Mathematical Programming. McGraw-Hill, 1970.

Clark, C. W., Mathematical Bioeconomics: The Optimal Management of Renewable Resources. Wiley, 1976.

Dorfman, R., "An Economic Interpretation of Optimal Control Theory," American Economic Review. December 1969, pp. 817-831.

Intriligator, M. P., Mathematical Optimization and Economic Theory. Prentice-Hall, 1971, Ch. 14, pp. 344-353.

Kamien, M. I. and N. O. Schwartz, Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management. North-Holland, 1981.

Knight, F. H., "Diminishing Returns from Investment," Journal of Political Economy 52 (March 1944).

Malliaris, A. G. and W. A. Brock, Stochastic Methods in Economics and Finance. North-Holland, 1982.

Varian, H. R., Microeconomic Analysis. W. W. Norton and Company, Inc., 1978.

Appendix A. Principles of Constrained Optimization

A. A Review of Optimization

One of the first lessons one learns in calculus is how to find an extreme value of a smooth function, $f(x)$, of one variable, x. The rules are easily visualized and remembered:

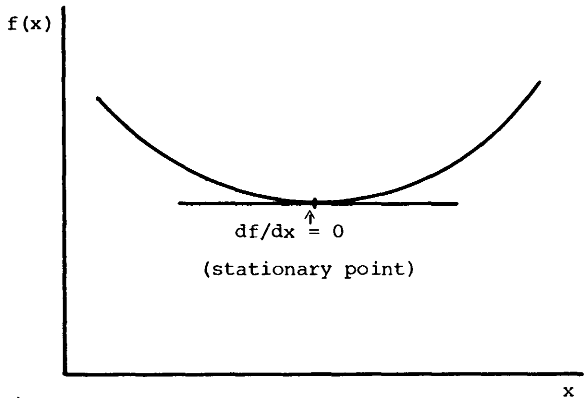
(a) When the derivative of f with respect to x equals zero, $df/dx = 0$, the function $f(x)$ has a (local) stationary point.

(b) If for all values of x the second derivative of $f(x)$, d^2f/dx^2, is positive, then $f(x)$ has a global minimum at the value of x where $df/dx = 0$. Such a function has a graph that looks like a valley for all x and is called strictly convex [Figure A1(a)].

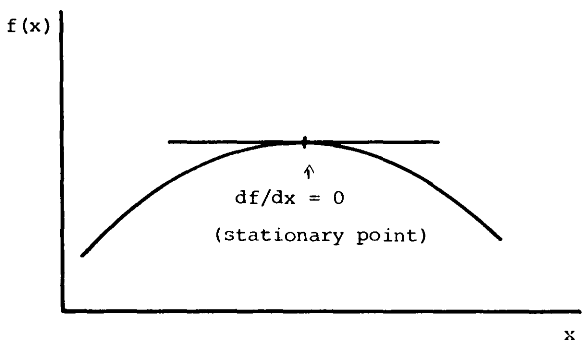
(c) If for all values of x the second derivative of $f(x)$ is negative, then $f(x)$ has a global maximum at the value of x where $df/dx = 0$. Such a function has a graph that looks like a hill for all x and is called strictly concave [Figure A1(b)].

(d) If the second derivative of $f(x)$ changes sign, then a local stationary value may be neither a local maximum nor a local minimum, much less a global maximum or minimum. Such a function is neither convex nor concave [Figure A1(c)]. If a function has a flat spot so that the second derivative equals zero for an interval of x, the function may be convex without being strictly convex or concave without being strictly concave; in this case a value of x giving a local stationary point may not be the only value of x to give a maximum or minimum [Figure A1(d), Figure A1(e)]. For this reason I work only with strictly convex or strictly concave functions in this report.

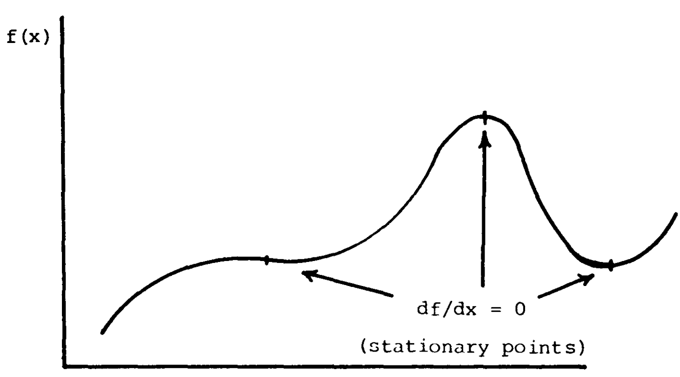
Figure A1 Extreme values of smooth functions of a single variable



A1(a) A strictly convex function

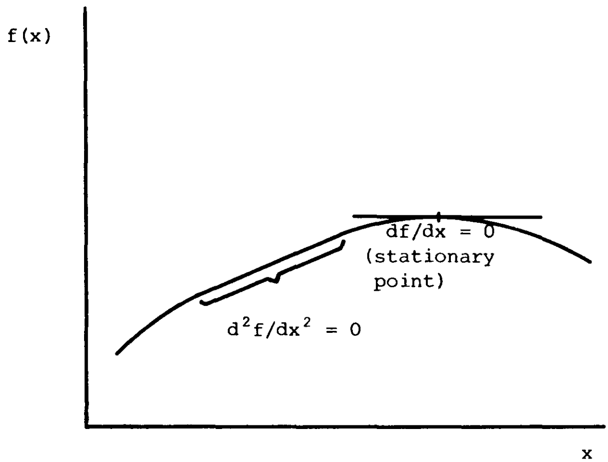


A1(b) A strictly concave function

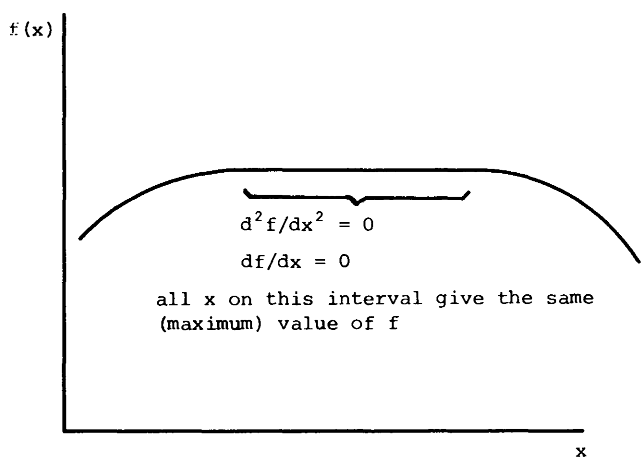


A1(c) A function that is neither convex nor concave

Figure A1 (continued)



A1(d) A function that is concave but not strictly concave but with a global maximum



A1(e) A function that is concave but not strictly concave and a maximum at many values of x

When we seek the extreme points of a smooth function, $f(x_1, x_2)$, of two variables, x_1 and x_2, the rules are very similar to those for the single variable problem. Recall that $\partial f/\partial x_1$ denotes the derivative of $f(x_1, x_2)$ with respect to x_1 while holding x_2 constant. Therefore, $\partial f/\partial x_1$ is the rate of change of f in the x_1 direction. For a function of several variables, the conditions on the second derivatives for a strictly convex or strictly concave function become more involved. But the picture of a smooth valley or hill remain valid in higher dimensions

(a) When the function f is strictly convex or strictly concave and the partial derivative of f with respect to x_1 and the partial derivative of f with respect to x_2 both equal zero, $\partial f/\partial x_1 = \partial f/\partial x_2 = 0$, the function f(x) has a maximum or minimum at the point (x_1, x_2).

(b) If the function $f(x_x, x_2)$ is strictly convex, the function has a global minimum at this point.

(c) If the function is strictly concave, the function has a global maximum at this point.

Since it is difficult to draw three-dimensional diagrams, the function of two variables is represented by contour lines (also called level lines). Along each contour line the function has a single value. Figure 2 illustrates this problem.

B. Maximization Subject to an Equality Constraint

Economists frequently discuss the problem of maximizing a utility or objective function subject to an income or expenditure constraint. In this type of constraint the total expenditure on x_1 and x_2 is required to be a given value. This gives a linear equality constraint such as

$$p_1x_1 + p_2x_2 = I. \qquad (1)$$

The problem is to attain the highest level of the objective function, $f(x_1, x_2)$, possible while remaining on the constraint line. This problem and the solution point are illustrated in Figure 3. Notice

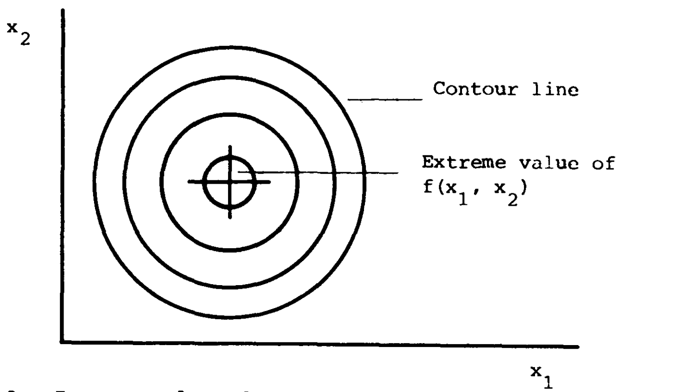


Figure 2. Extreme value of a smooth function of two variables

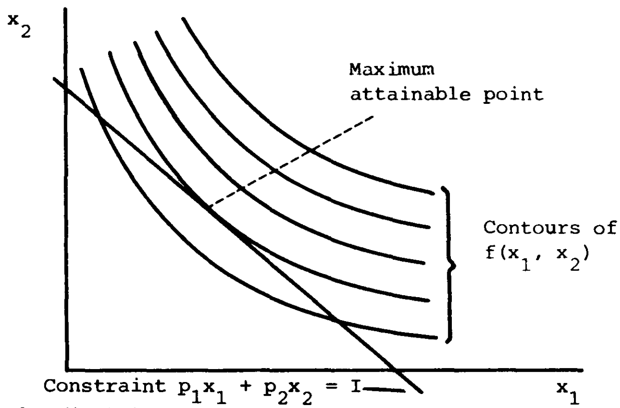


Figure 3. Maximizing an objective function subject to a linear equality constraint

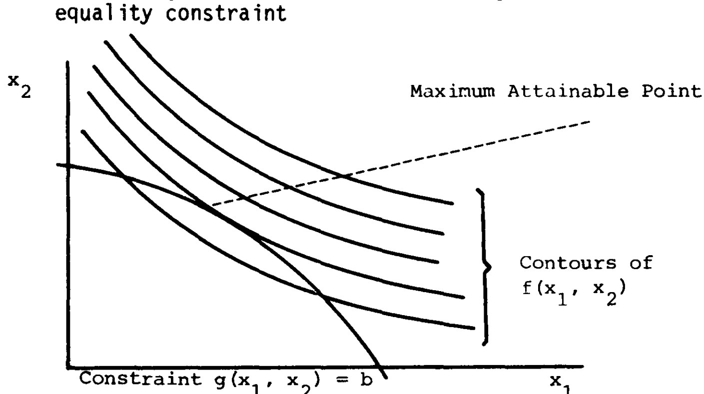


Figure 4. Maximizing an objective function subject to a concave equality constraint

that in the maximization problem the contours represent a "smooth hill sloping up toward the Northeast." To avoid complications, this report deals only with objective functions $f(x_1, x_2)$ that are strictly concave. But the contours each show x_2 as a strictly convex function of x_1 (or x_1 as a strictly convex function of x_2).

As shown in Figure 4, the constraint on attainment of the objective may be a nonlinear function, such as a production function, of the variables x_1, x_2. The nonlinear equality constraint is written in the form

$$g(x_1, x_2) = b. \tag{2}$$

It is convenient to think of the x's as activities and the g as a resource. The activities may be either consumption activities (e.g., consume x_1 = bread, x_2 = butter) or production activities (e.g., produce x_1 = bread, x_2 = butter). Both activities use the resource g, which is limited in quantity to b. If the amount of the resource available is reduced, the constraint contour would be closer to the origin but would have a similar shape. To avoid complications this report deals only with constraint functions that are strictly convex. But the contours each show x_2 as a strictly concave function of x_1 (or x_1 as a strictly concave function of x_2). The familiar marginal conditions determine the maximum of the objective if the contour of the constraint function is concave and the contours of the objective function are strictly convex.

More formally the problem of maximization subject to an equality constraint is

$$\begin{aligned} &\text{Maximize } f(x_1, x_2) \\ &\text{w.r.t. } x_1, x_2 \\ &\text{subject to } g(x_1, x_2) = b. \end{aligned} \tag{3}$$

This notation means to maximize the function $f(x_1, x_2)$ with respect to (w.r.t.) x_1 and x_2. By use of the device of a Lagrange multiplier, λ, the problem (3) is changed from a constrained maximization problem to an unconstrained maximization problem. Thus, the solution to (3) is found if we maximize the Lagrangian function:

$$\text{Maximize } L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda(b - g(x_1, x_2))$$
$$\text{w.r.t. } x_1, x_2, \lambda. \qquad (4)$$

The Lagrange multiplier λ is a finite constant and with the shapes I have assumed for the objective and constraint functions, it is non-negative when we are careful to express the constraint in the form $(b - g(x_1, x_2))$. If λ $(g - b)$ is added to f, then λ would turn out to have a nonpositive value.

The necessary conditions for solution of problem (4) are

$$\begin{aligned} &(a) \quad \partial L/\partial x_1 = \partial f/\partial x_1 - \lambda\, \partial g/\partial x_1 = 0 \\ &(b) \quad \partial L/\partial x_2 = \partial f/\partial x_2 - \lambda\, \partial g/\partial x_2 = 0 \\ &(c) \quad \partial L/\partial \lambda = b - g = 0. \end{aligned} \qquad (5)$$

Denote the optimum values, given by the solution of equations (5), as x_1^*, x_2^*, and λ^*. The reason for the necessary condition for a maximum with respect to x_2 is as follows. Because we can increase <u>or</u> decrease x_2, at the optimum values of x_1 and λ_1

(a) if L is decreasing in the x_2 direction, i.e., $\frac{\partial L}{\partial x_2} < 0$, then we would decrease x_2 to maximize L.

(b) if L is increasing in the x_2 direction, i.e., $\frac{\partial L}{\partial x_2} > 0$, then we would increase x_2 to maximize L.

Therefore, at the optimal values of x_1, x_2 and λ, the value of L must not be changing with x_2, i.e.,

$$\frac{\partial L}{\partial x_2} = 0.$$

The reasoning for x_1 is exactly the same except that x_2 is held at the optimum value. It should be clear to the reader that when the constraint is satisfied, g = b, and $L(x_1, x_2, \lambda)$ has the same value as $f(x_1, x_2)$. Therefore, since the optimal values, x_1^*, x_2^*, λ^* satisfy equation 5(c), we must have L = f at the optimum values, x_1^*, x_2^*, λ^*.

Two interrelated questions have puzzled students before they before they became very familiar with the theory of mathematical programming. (1) How do you interpret the optimal value of the multiplier, λ^*? (2) How do you explain what determines the optimal value of the multiplier?

The Lagrange multiplier, λ, is often called the shadow price of the resource g.

> THE OPTIMAL VALUE λ^* OF THE MULTIPLIER TELLS HOW MUCH A MARGINAL INCREASE IN THE RESOURCE g IS WORTH.

If $f(x_1, x_2)$ is a utility function, then λ^* is the marginal utility of g*, if $f(x_1, x_2)$ is a production function,[1] then λ^* is the marginal product of g.

Figure 5 expands Figure 4 to demonstrate the effect of increasing the amount of resource available by the amount Δ. From Figure 5 it is clear that (before we pass the global maximum) increasing the amount of resource available will increase the attainable value of the objective function. Therefore, in Figure 6 the attainable value is shown as an increasing function of the amount of resource used.

In Figure 6 look at the curve labeled $L|b$, which means L given b. When g is less than b, the curve $L|b$ is above the curve f*. When g is greater than b, the curve $L|b$ is below the curve f*. When g equals b, the curve $L|b$ crosses the curve f* and the curve $L|b$ is at its maximum. (Be careful to notice that in Figure 6 the optimum value of the multiplier λ^* is assumed at all times.) To see why the curve $L|b$ is this way, look back at equation (4). From equation (4) we see that if the amount of the resource used were less than the amount available, b, then the Lagrangian, L, would be greater than (or equal to) the value of the objective f because the optimal value of the multiplier, λ^*, is a nonnegative number. Conversely, if the amount of resource used were greater than the amount available, then L would be less than or equal to f. Also, satisfying the conditions of equation (5) assures that L is a maximum when exactly the available amount of the resource is used, i.e., g = b.

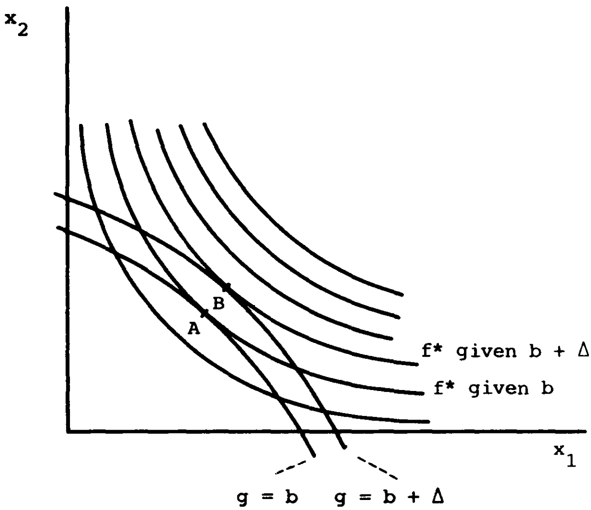


Figure 5. Effect of increasing the amount of resource available

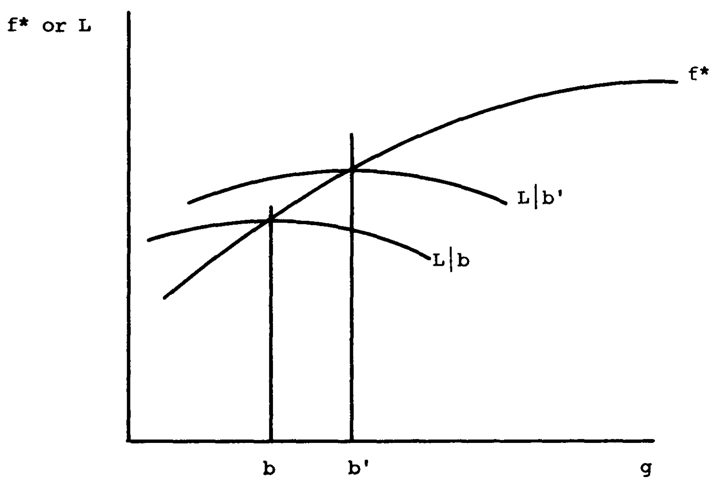


Figure 6. Attainable objectives as a function of resource used

THE OPTIMAL VALUE OF THE MULTIPLIER, λ^*, IS THE VALUE THAT WOULD CAUSE ONE TRYING TO MAXIMIZE THE LAGRANGIAN, L, TO DEMAND EXACTLY THE AVAILABLE QUANTITY OF THE RESOURCE g IF THE RESOURCE COULD BE BOUGHT AND SOLD AT THE PRICE λ^* PER UNIT OF g.

If the available amount of the resource g increases to $b' = b + \Delta$, we move to a higher path of L, labeled $L|b'$ in Figure 6. With b' available, a higher level of the objective is obtained. A new value of λ will now assure that $L|b'$ reaches its maximum at b'.

The determination of the optimal value of the multiplier λ^* can also be expressed in terms of supply and demand. Think of λ as the price on a demand schedule for the resource g. This is justified by noting that, as shown above, λ is the increment in the objective function per unit of increment in the resource available.
Thus, in consumption studies, λ is the marginal utility per unit of resource g; in production studies, λ is the value of the marginal product per unit of resource g. The optimal value of λ is the value that causes exactly the available quantity of the resource to be demanded. See Figure 7. In this way it is natural to call λ the shadow price of the resource. The "shadow" adjective is used because this price does not appear in a market and is not actually paid.

THE OPTIMAL VALUE OF THE MULTIPLIER, λ^*, IS THE AMOUNT THAT THE OPTIMIZER WOULD BE WILLING TO PAY FOR AN ADDITIONAL UNIT OF THE RESOURCE g. HENCE, λ^* IS CALLED THE SHADOW PRICE OF THE RESOURCE g.

C. Limits on Activities (Variables x_1, x_2)

1. Nonlinear Objective Functions. Now consider a slightly more difficult problem. Suppose that activity x might be driven to a lower limit ℓ_1 before the maximum sought for the problem in equation (3) is reached. The problem is then

$$\begin{aligned} &\text{Maximize } f(x_1, x_2) \\ &\text{w.r.t. } (x_1, x_2) \\ &\text{subject to } g(x_1, x_2) = b \,. \\ &\qquad\qquad x_1 \geq \ell_1 \end{aligned} \tag{6}$$

The Lagrangian for the problem in equation (6) is very similar to equation (4)

$$\text{Maximize } L(x_1, x_2\ \lambda) = f(x_1, x_2) + \lambda(b - g(x_1, x_2))$$
$$\text{w.r.t. } x_1, x_2, \lambda$$
$$\text{subject to } x_1 \geq \ell_1, \qquad (7)$$

but in maximizing $L(x_1, x_2, \lambda)$, some of the necessary conditions are different if x_1 is driven to ℓ_1. The resource constraint still must be satisfied so $b - g = 0$.

Suppose that the optimum value of activity x_1 is the lower limit, i.e., $x_1{}^* = \ell_1$. This situation is plotted in Figure 8. Now, proceed very carefully in figuring out what the conditions on the partial derivatives of L must be in this situation.

The reasoning proceeds as follows:

(a) We <u>cannot</u> decrease x_1 below ℓ_1, thus at the optimum we <u>can</u> have $\frac{\partial L}{\partial x_1} \leq 0$. The condition $\frac{\partial L}{\partial x_1} < 0$ tells us that we could increase L by decreasing x_1, but since x_1 cannot be below ℓ_1, we are stuck at ℓ_1 for the optimum value of x_1. The equality part of the $\leq$ sign allows for the possibility that the limit is the unconstrained optimum.

(b) We <u>can</u> increase x_1 above ℓ_1, thus at the optimum we <u>cannot</u> have $\frac{\partial L}{\partial x_1} > 0$. The condition $\frac{\partial L}{\partial x_1} > 0$ tells us that we could increase L by increasing x_1. If that were so, then ℓ_1 would not be the optimum value of x_1.

(c) At any value of $x_1{}^*$ other then ℓ_1, it is necessary that at the optimum $\frac{\partial L}{\partial x_1} = 0$. Either $x_1{}^* = \ell_1$ or $\frac{\partial L}{\partial x_1} = 0$ at the optimum. This condition can be written $(x_1{}^* - \ell_1)\frac{\partial L}{\partial x_1} = 0$. Therefore, the necessary conditions for

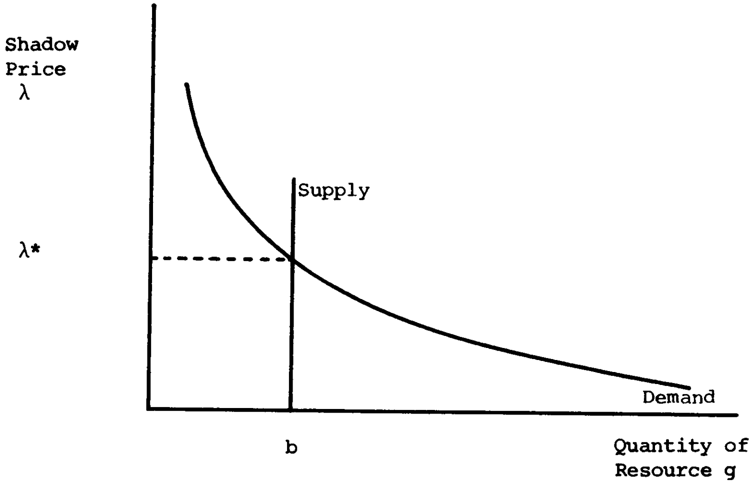


Figure 7. Demand-supply determination of the shadow price of resource g

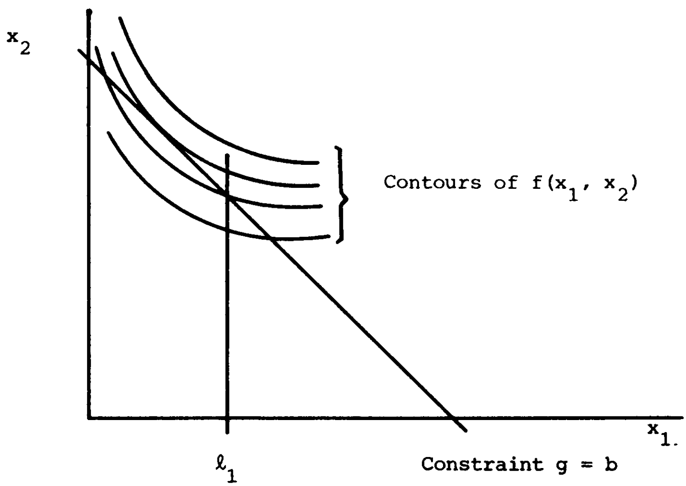


Figure 8. Maximization subject to lower limit on x_1

a maximum with respect to x_1 is, at the optimum,

$$\frac{\partial L}{\partial x_1} \leq 0 \text{ AND } (x_1^* - \ell_1)\frac{\partial L}{\partial x_1} = 0,\ x_1^* \geq \ell_1.$$

The necessary conditions for solution of problem (6) are

$$\text{(a)}\quad \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda\frac{\partial g}{\partial x_1} \leq 0; \qquad (x_1^* - \ell_1)\frac{\partial L}{\partial x_1} = 0$$

$$x_1 \geq \ell_1$$

$$\text{(b)}\quad \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda\frac{\partial g}{\partial x_2} = 0 \qquad (8)$$

$$\text{(c)}\quad \frac{\partial L}{\partial \lambda} = b - g(x_1, x_2) = 0 .$$

In the maximization problem any variable subject to a lower bound gives a necessary condition such as (8a).

Now to demonstrate the effect of an upper bound, suppose that x_1 has no bounds, but x_2 is subject to an upper bound u_2. The problem of equation (3) is then

$$\begin{aligned} &\text{Maximize } f(x_1, x_2) \\ &\text{w.r.t. } x_1, x_2 \\ &\text{subject to } g(x_1, x_2) = b \\ &\qquad\qquad x_2 \leq u_2. \end{aligned} \qquad (9)$$

The Lagrangian is again as in equation (4) but now is subject to $x_2 \leq u_2$. Reasoning exactly similar to that above shows that the necessary conditions for solution of problem (9) are

$$\text{(a)}\quad \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0$$

(b) $\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda\frac{\partial g}{\partial x_2} \geq 0;$ $\qquad (x_2^* - u_2)\frac{\partial L}{\partial x_2}$ (10)

$x_2^* \leq u_2$

(c) $b - g(x_1, x_2) = 0.$

SOLUTIONS SUCH AS THOSE THAT DRIVE A VARIABLE TO ITS LOWER OR UPPER LIMIT ARE CALLED CORNER SOLUTIONS. FOR CORNER SOLUTIONS THE NECESSARY CONDITIONS (8a or 10b) ALLOW FOR INEQUALITIES

Suppose the problem is to minimize a convex function $k(v_1, v_2)$, such as a cost function, subject to an equality constraint on a resource and a lower limit on both v_1 and v_2.

Minimize $k(v_1, v_2)$

w.r.t. v_1, v_2

subject to $h(v_1, v_2) = c$ (11)

$v_1 \geq \ell_1$

$v_2 \geq \ell_2.$

The Lagrangian problem is

Minimize $L(v_1, v_2, \lambda) = k(v_1, v_2) + \mu(c - h(v_1, v_2))$

w.r.t. v_1, v_2, μ

subject to $v_1 \geq \ell_1 \qquad v_2 \geq \ell_2.$ (12)

A different symbol is used for the Lagrange multiplier to distinguish between maximization problems and minimization problems. With the same type of reasoning as for problem (6), we see that the necessary conditions for solution of minimization problem (12) is

$$\frac{\partial L}{\partial v_1} = \frac{\partial k}{\partial v_1} - \frac{\partial h}{\partial v_1} \geq 0; \qquad (v_1^* - \ell_1)\frac{\partial L}{\partial v_1} = 0$$

$$v_1 \geq \ell_1$$

$$\frac{\partial L}{\partial v_2} = \frac{\partial k}{\partial v_2} - \frac{\partial h}{\partial v_2} \geq 0; \qquad (v_2^* - \ell_2)\frac{\partial L}{\partial v_2} = 0 \tag{13}$$

$$\frac{\partial L}{\partial \mu} = C - h(v_1, v_2) = 0. \qquad v_2 \geq \ell_2$$

The value of the Lagrange multiplier at the optimal values (v_1^*, v_2^*, μ^*) still gives the change in the objective function $k(v_1, v_2)$ per unit increase in the resource h. However, since the objective is to reduce costs to its minimum, a negative value of μ indicates a positive value of the resource in attaining the objective of reduced cost.

Some analysts are more accustomed to working with minimization problems than with maximization problems. For minimization problems it is usually preferable to write the Lagrangian function in the form $L = k(v_1, v_2) - \mu(c - h(v_1, v_2))$. Written in this way the optimum value of μ is the negative of the change in the objective function and hence a positive value of μ then indicates a positive value of the resource h in attaining the objective of reduced cost.

When the Lagrangian is written with $\lambda(b-g)$ in a maximization problem, a negative value of λ indicates a positive value of the resource g in attaining the objective of increased product or utility.

2. Linear Objective Functions and Linear Constraints. If the objective function is a sloping plane in three dimensions (x_1, x_2, f), the contours of f are straight lines. If the contours of the objective function have exactly the same slope as the constraint, all points along the constraint are equally good solutions. When there is one constraint that is a straight line with a slope different from the slope of the objective function, one of the activities x_1, x_2 must be limited to provide a definite solution. Such a problem would be

$$\begin{aligned}&\text{Maximize } f(x_1, x_2) = c_1x_1 + c_2x_2\\&\text{w.r.t. } x_1, x_2\\&\text{subject to } a_1x_1 + a_2x_2 = b \qquad (14)\\&\qquad\qquad x_1 \geq 0.\end{aligned}$$

Problem (14) is presented graphically in Figure 9. This figure shows that, unless the contours have exactly the same slope as the constraint, the solution must be a corner solution.

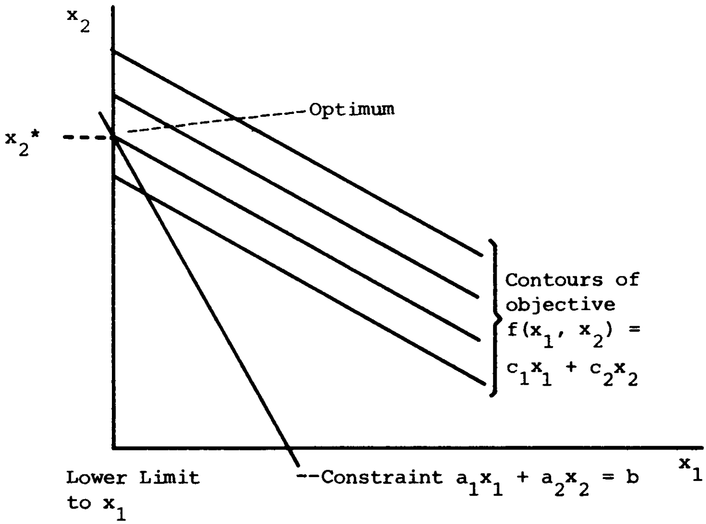


Figure 9. Maximization of a linear objective function subject to a linear equality constraint

The Lagrangian function problem corresponding to problem (14) is

$$\begin{aligned}&\text{Maximize } L(x_1, x_2, \lambda) = c_1x_1 + c_2x_2 + \lambda(b - a_1x_1 - a_2x_2)\\&\text{w.r.t. } x_1x_2\\&\text{subject to } x_1 \geq 0. \qquad (15)\end{aligned}$$

The necessary conditions for solution of problem (15) are as in equation (8).

$$\text{(a)} \quad \frac{\partial L}{\partial x_1} = c_1 - a_1\lambda \leq 0 \qquad (x_1^*)(c_1 - a_1\lambda^*) = 0$$

$$\text{(b)} \quad \frac{\partial L}{\partial x_2} = c_2 - a_2\lambda = 0 \qquad x_1 \geq 0$$

$$\text{(c)} \quad b - a_1x_1 - a_2x_2 = 0. \tag{16}$$

Since the optimal value of x_1 is zero, 16(a) may be an inequality. From 16(b) we find that the shadow price of the resource is $\lambda^* = c_2/a_2$. Interpret $\lambda^* = c_2/a_2$ as follows:

(a) From the objective function in (14) notice that the objective is increased c_2 units for each unit increase in activity (variable) x_2.

(b) From the constraint 16(c) notice that, since $x_1^* = 0$, for each one-unit increase in the quantity of resource available, b, the activity x_2 may be increased $1/a_2$ units so that the optimum value of x_2 is $x_2^* = b/a_2$

(c) Hence, the marginal value of the resource, λ^*, is [the increase in the objective per unit increase in activity x_2] multiplied by
[the increase in units of x_2 per unit increase in resource]

$$\lambda^* = c_2 \cdot 1/a_2.$$

Do not introduce another equality constraint into the problem with two activities x_1, x_2 because two equations determine the solution for two variables: there would be no optimization problem left after satisfying the constraints. Also, the two constraint equations would have to be consistent with each other and would have to yield a solution in the allowed range of the activities x_1, x_2. However, to demonstrate how easily the two activity problems may be generalized, a problem for more than two variables with more than one equality constraint is shown in the following paragraphs.

3. More Activities and Constraints. In this problem there are n activities x_1, x_2 . . ., x_n, and the objective is to maximize the value of the function $f(x_1, x_2, \ldots x_n)$ subject to equality

constraints on the use of m resources and lower bounds on each of the activities. The main complication of this extension is the requirement that the number of resource equality constraints be less than (or equal to) the number of activities, $m \leq n$. If there are n different equality constraints, $m = n$, that are not contradictory and that have a solution in the allowed range of activities, there is only one point that satisfies the constraints and no optimization problem is involved. Assuming that $m < n$ the problem is

$$\begin{aligned} &\text{Maximize } f(x_1, x_2, \ldots, x_n) \\ &\text{w.r.t. } x_1, x_2, \ldots, x_n \\ &\text{subject to } g_1(x_1, x_2, \ldots, x_n) = b_1 \\ &\qquad g_2(x_1, x_2, \ldots, x_n) = b_2 \\ &\qquad g_m(x_1, x_2, \ldots, x_n) = b_m \\ &\qquad x_1 \geq \ell_1, x_2 \geq \ell_2, \ldots, x_n \geq \ell_n. \end{aligned} \tag{17}$$

The first order conditions are straightforward generalizations of (8). With

$$L = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i(b_i - g_i(x_1, x_2, \ldots, x_n)).$$

The necessary conditions are

$$\begin{aligned} &(a) \quad \frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} \leq 0; \quad (x_j^* - \ell_j) \frac{\partial L}{\partial x_j} = 0 \\ &\qquad x_j \geq \ell_j \qquad\qquad j = 1, 2, \ldots n \\ &(b) \quad \frac{\partial L}{\partial \lambda_j} = b_i - g_i = 0 \qquad \text{for } i = 1, 2, \ldots, m. \end{aligned} \tag{18}$$

These conditions (18) are interpreted as before. The optimal shadow price λ_i^* is the marginal value of the i^{th} resource at the optimum. The additional fact that should be remembered is that a change in the available quantity of one resource may change not only its own shadow price but also the shadow prices of other resources as well as the optimum values of any or all activities.

To save space we will now frequently use the notation $\underline{x}$ to stand for all of the activities $(x_1, x_2, \ldots, x_n)$, and $\underline{\lambda}$ to stand for all of the shadow prices $(\lambda_1, \lambda_2, \ldots \lambda_m)$. Remember also the meaning of the summation notation $\sum_{i=1}^{n} y_i = y_1 + y_2 + \ldots + y_n$.

D. Inequality Constraints on Resources

In problems with several resource constraints, it usually is not desirable to require that the constraints be satisfied as equalities. A firm may have a maximum of ten machines that can be used in production, but the optimal mix of production activities frequently may leave one or more of the machines unemployed. An inequality constraint gives only the limit on a resource, it does not require that all of a resource be used.

Begin by considering a problem with two activities and two resource inequality constraints. This problem is diagrammed in Figure 10 in which only the points between the origin and the resource constraints are attainable. Notice that since the constraints are inequalities the two constraints do not completely determine the solution for the value of the two activities. In Figure 10 the optimum point is point A. Since A is inside of the boundary for the first constraint, $g_1(x_1, x_2) < b_1$, this means that some of the available resource g_1 is unemployed. If there were no constraint on resource g_2, the optimum would be at point B where all of resource g_1 would be employed.

Stated formally, the problem depicted in Figure 10 is

$$\begin{aligned} &\text{Maximize } f(x_1, x_2) \\ &\text{w.r.t. } x_1, x_2 \\ &\text{subject to } g_1(x_1, x_2) \leq b_1 \\ &\qquad\qquad\;\; g_2(x_1, x_2) \leq b_2. \end{aligned} \tag{19}$$

It is wise to follow Baumol's (1977, p. 160) suggestion and for a maximization problem arrange the inequalities in the form $b_i - g_i \geq 0$ so that the Lagrangian will be written in the same form as in equation (14) and yield nonnegative multipliers. To convert problem (19) into a problem that you know how to solve, subtract

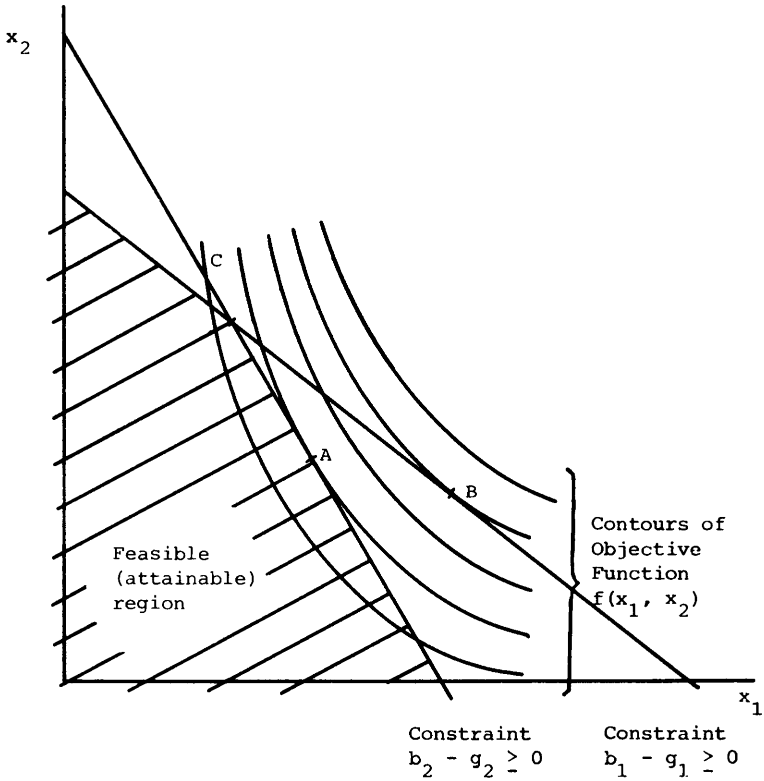


Figure 10. Maximization with respect to two activities subject to two inequality resource constraints

nonnegative "slack" variables $S_i \geq 0$ from each $b_i - g_i$. Problem (19) can then be written

$$\begin{aligned} &\text{Maximize } f(x_1, x_2) \\ &\text{w.r.t. } x_1, x_2, S_1, S_2 \\ &\text{subject to } b_1 - g_1(x_1, x_2) - S_1 = 0 \\ &\qquad b_2 - g_2(x_1, x_2) - S_2 = 0 \\ &\qquad S_1 \geq 0,\ S_2 \geq 0. \end{aligned} \tag{20}$$

Adding two new "activities," S_1 and S_2, to the problem gets rid of the inequalities in the constraints. It does not matter that the S's are not in the objective function; those partial derivatives are just zero, e.g., $\frac{\partial f}{\partial S_1} = 0$. There is a side benefit from having resource constraints as inequalities; now the number of constraints is less than the number of activities in the objective function because a new activity is added for each constraint and this assures that there are more activities than there are equality constraints in the problem.

Now solve problem (2) in the same way as problem (17). Form the Lagrangian function

$$\begin{aligned} L(x_1, x_2, \lambda_1, \lambda_2, S_1, S_2) &= f(x_1, x_2) \\ &+ \lambda_1(b_1 - g_1(x_1, x_2) - S_1) \\ &+ \lambda_2(b_2 - g_2(x_1, x_2) - S_2) \end{aligned}$$

subject to $S_1 \geq 0$ $S_2 > 0$

and write the necessary conditions. There are no limits on the activities x_1, x_2, but the slack variables must be nonnegative. The necessary conditions are

(a) $$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} - \lambda_2 \frac{\partial g_2}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda_1 \frac{\partial g_1}{\partial x_2} - \lambda_2 \frac{\partial g_2}{\partial x_2} = 0$$

$$\text{(b)} \quad \frac{\partial L}{\partial S_1} = -\lambda_1 \leq 0 \qquad S_1{}^* \frac{\partial L}{\partial S_1} = 0 \qquad S_1 \geq 0$$

$$\frac{\partial L}{\partial S_2} = -\lambda_2 \leq 0 \qquad S_2{}^* \frac{\partial L}{\partial S_2} = 0 \qquad S_2 \geq 0 \qquad (21)$$

$$\text{(c)} \quad \frac{\partial L}{\partial \lambda_1} = b_1 - g_1 - S_1 = 0$$

$$\frac{\partial L}{\partial \lambda_2} = b_2 - g_2 - S_2 = 0.$$

The conditions (21) are more elegant and more easily interpreted if a few algebraic manipulations are performed. Multiply the first half of (21b) by -1 and substitute into the second half of (21b) to get the conditions in the form

$$\lambda_i \geq 0 \qquad S_i{}^*\lambda_i{}^* = 0 \qquad i - 1, 2.$$

Condition (21c) shows that if $S_i{}^* = 0$, then $b_i - g(\underline{x}^*) = 0$, so substitute $b_i - g_i(\underline{x}^*)$ for $S_i{}^*$ in conditions (21b). Also, write condition (21c) in the original form of the constraints. Therefore, the necessary conditions for the solution of problem (20) are

$$\text{(a)} \quad \frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \qquad j = 1, 2$$

$$\text{(b)} \quad \lambda_i \geq 0 \qquad \lambda_i{}^*(b_i - g_i(\underline{x}^*)) = 0 \qquad i = 1, 2 \qquad (22)$$

$$\text{(c)} \quad g_i \leq b_i \,. \qquad i = 1, 2$$

THE SHADOW PRICES ARE NONNEGATIVE. THEY ARE ZERO IF ANY OF THE RESOURCE IS UNEMPLOYED IN THE OPTIMAL SOLUTION.

The cook wouldn't pay for additional pans if he already had more than enough pans to bake all of the cakes he wanted to bake. With the conditions shown in Figure 10, $\lambda_1 = 0$ since at A some of resource 1 is not used, $g_1 < b_1$. The solution at point A is, therefore, as if the constraint on resource 1 did not exist.

With a convex feasible region, which gives a concave boundary as in Figure 10, and a concave objective function, which gives convex contours as in Figure 10,

THE CONDITIONS OF EQUATION 22 CAN BE SATISFIED SIMULTANEOUSLY ONLY AT THE MAXIMUM POINT.

There are two facts that the reader should note:

(1) The "slack" variables, S's, work like catalysts; they help the process but do not appear in the final product.

(2) With two inequality constraints, as in Figure 10, there is the possibility of a corner solution at point C even though there are no limits on the range of the activities x_1 and x_2 other than the resource constraints. If the objective function is such that point C is the optimum feasible choice, the conditions of equation (22) still hold, but the highest contour attained may not be tangent to the frontier of either constraint. This possibility is diagrammed in Figure 11.

Before proceeding with additional complications to this problem, a numerical example may help. The fact that the slope of the objective function contours need not be tangent to either constraint frontier seems obvious from Figure 11. However, this possibility is not so obvious in the mathematical equations (22) for the necessary conditions. In the following subsection a numerical example demonstrates this point. It is not necessary for the reader to follow through this example in order to proceed with the main discussion; therefore, the reader may wish to skip the following section for the present.

E. An Example of Maximization Subject to Inequality Constraints

For an example of constrained maximization, turn to one of economists' favorite literary characters. Robinson Crusoe has only two activities by which he survives: production of food, F, and clothing, C. His utility function for F and C is $U = F^{2/3}C^{1/3}$. He produces food and clothing with only two resources: land, L, and his own effort, E. During the production season he can work a maximum of 150 days and has 100 ares of land for cultivation (100 ares = one hectare = 2.47 acres). Each unit of food produced requires a day of effort and an are of land. Each unit of clothing

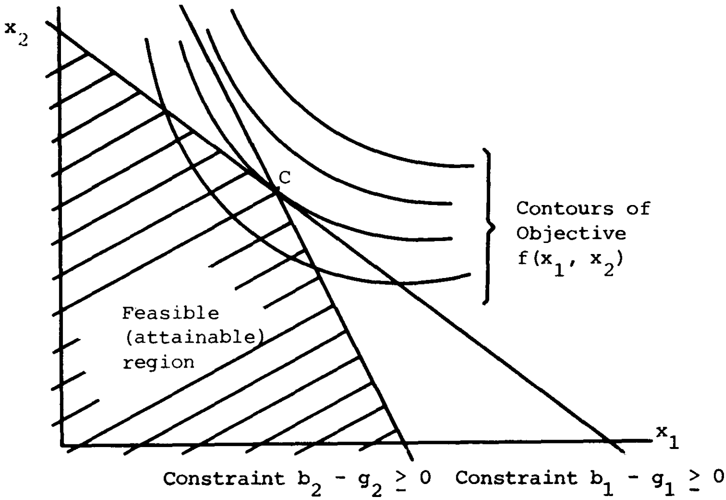


Figure 11. Maximization with two activities and two inequality constraints leading to a corner solution

produced requires three days effort and an are of land.

Rob's constraint on effort is 150 days, so $F + 3C \leq 150$. If he put all of his effort into producing food, he could produce at most 150 units with unlimited land; if he put all of his effort into producing clothes, he would produce at most 50 units with unlimited land. His constraint on land is 100 ares, so $F + C \leq 100$. If there were no limit on his effort, he could produce either 100 units of food or 100 units of clothing.

Ignoring for now the requirement that both food and clothing must be produced in nonnegative amounts, Rob's problem is

$$\begin{aligned} &\text{Maximizing } U(F, C) = F^{2/3}C^{1/3} \\ &\text{w.r.t. } F, C \\ &\text{subject to } F + 3C \leq 150 \qquad \text{(effort constraint)} \\ &\qquad\qquad\quad F + C \leq 100 \qquad \text{(acreage constraint)} \end{aligned}$$

with Lagrangian

$$L(F, C, \lambda_1, \lambda_2, S_1, S_2) = F^{2/3}C^{1/3} + \lambda_E(150 - F - 3C - S_1) + \lambda_A(100 - F - C - S_2).$$

Following equation (22), the necessary conditions for solving this problem are

$$\text{(a)} \quad \frac{2}{3} F^{-1/3}C^{1/3} - \lambda_E - \lambda_A = 0$$

$$\frac{1}{3} F^{2/3}C^{-2/3} - 3\lambda_E - \lambda_A = 0$$

$$\text{(b)} \quad \lambda_E \geq 0 \qquad \lambda_E{*}(150 - F - 3C) = 0 \qquad (24)$$

$$\lambda_A \geq 0 \qquad \lambda_A{}^{*}(100 - F - C) = 0$$

$$\text{(c)} \quad F + 3C \leq 150$$

$$F + C \leq 100.$$

Here $\lambda_E{*}$ is the optimal shadow price of a day of effort and $\lambda_A{*}$ is the optimal shadow price of an acre of land.

This problem is drafted rather precisely in Figure 12. From this figure it is seen that the optimum production is 75 units of food and 25 units of clothing. All resources are fully utilized, since the maximum utility (52.00 units of utility) is attained at the junction of the frontiers of both constraints. Substituting F = 75, C = 25 into (24a) and solving yields the values

$$\lambda_E{*} = \frac{1}{6}(\frac{1}{3})^{1/3} \qquad \lambda_A{*} = \frac{1}{2}(\frac{1}{3})^{1/3} .$$

The shadow price of effort relative to that of land is $\lambda_E{*}/\lambda_a{*} = 1/3$. The relative price of effort with respect to that of land is determined despite the fixed proportions of resource used in production. As emphasized by Friedman (1976, p. 174)

> SUBSTITUTION IN CONSUMPTION SUBSTITUTES FOR SUBSTITUTION IN PRODUCTION IN DETERMINING THE RELATIVE PRICES OF RESOURCES.

We are assured that the point F = 75, C = 25 is the maximum attainable because the slope of the contour of the objective function at this point is between the slopes of the constraints that converge at this point. The slope of the utility function at this point (75, 25) is obtained from the implicit differentiation rule

$$\frac{dC}{dF}\Big|_u = - \frac{\frac{\partial U}{\partial F}}{\frac{\partial U}{\partial C}} = -2\frac{C}{F} = - \frac{2}{3} ,$$

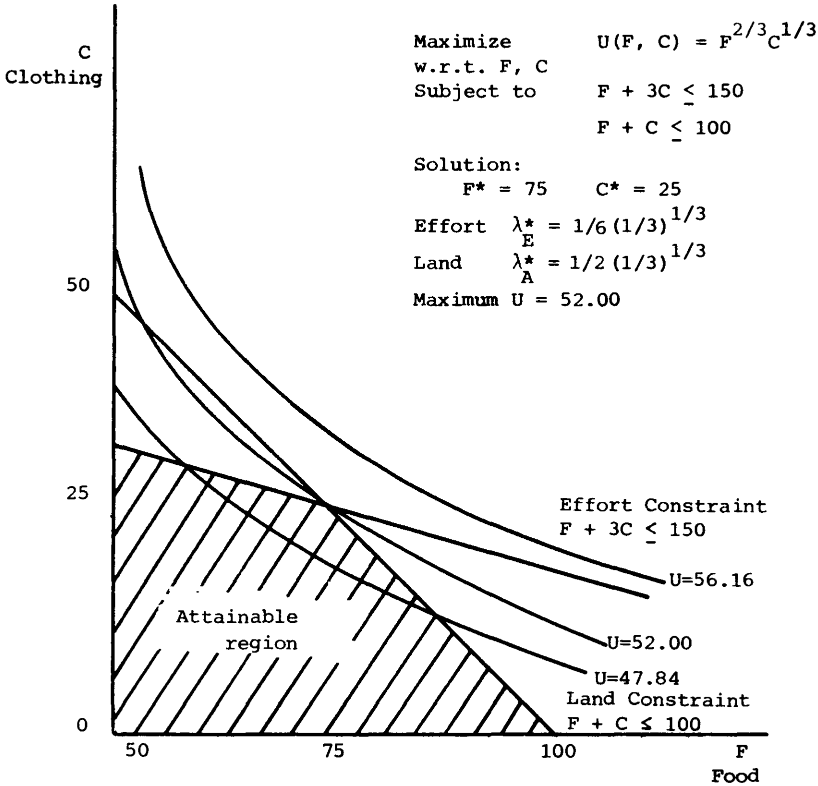


Figure 12. Example of maximization subject to inequality constraints

where the notation $\frac{dC}{dF}|_u$ means the change of C with respect to F along the function U(F, C).

The slope of the land constraint is $\frac{dC}{dF} = -1$ and the slope of the effort constraint is $\frac{dC}{dF} = -\frac{1}{3}$. If, at the point (75, 25), the slope $\frac{dC}{dF}|_u$ were less than -1 (greater than 1 in magnitude), then we would know we should move southeast along the land constraint to find the optimum. Conversely, if the slope $\frac{dC}{dF}|_u$ were greater than -1/3 (less than 1/3 in magnitude) we would know we should move northwest along the effort constraint.

A POSITIVE SHADOW PRICE ON A RESOURCE SAYS THAT WE SHOULD INCREASE THE USE OF THAT RESOURCE IF WE ARE NOT STOPPED BY THE CONSTRAINT ON THAT RESOURCE.

Similarly, a negative shadow price tells us to reduce the use of that resource in order to achieve a maximum. BUT REMEMBER: having these "correct" signs on the multipliers depends upon setting up the constraints as we have done. Recall that in a maximization problem the shadow price gives the rate at which the objective function would increase with increased supply of the constraining resource.

To assure yourself that conditions (24) would not be satisfied by some other point, use (24a) to calculate the values of λ_1 and λ_2 at the point (80, 20), which also lies on the frontier of the attainable region. At (80, 20), the λ_1, λ_2 values would be determined by

$$\frac{2}{3}(80)^{-1/3}20^{1/3} - \lambda_1 - \lambda_2 = 0$$

$$\frac{1}{3}(80)^{2/3}20^{-2/3} - 3\lambda_1 - \lambda_2 = 0.$$

These equations imply that $\lambda_1 = \lambda_2 = \frac{1}{3}(\frac{1}{4})^{1/3}$. However, this solution is inconsistent with condition (24B). Since the point (80, 20) is inside the effort constraint, $150 - F - 3C = 150 - 80 - 60 = 10 > 0$, which the first equation of (24b) tells us requires $\lambda_1^* = 0$. Therefore, the conditions of equations (24) are not satisfied at any point except (75, 25).

F. More Variables and Constraints

Problem (19) is easily generalized to contain more activities, more resource constraints and limits on the scope of the activities. For a more extensive development, see Lancaster (1968). First consider the maximization problem

$$\begin{aligned} &\text{Maximize } f(x_1, x_2, \ldots x_n) \\ &\text{w.r.t. } x_1, x_2, \ldots, x_n \\ &\text{subject to } g_i(x_i, x_2, \ldots, x_n) \leq b_i \quad i = 1, 2, \ldots m \\ &\qquad x_j \geq \ell_j \quad j = 1, 2, \ldots n. \end{aligned} \tag{25}$$

Following the convention of writing the resource constraints for a maximization problem in the form $b_i - g_i \geq 0$ and adding a slack variable for each resource inequality, the Lagrangian problem is

$$\begin{aligned} &\text{Maximize } L(\underline{x}, \underline{S}, \underline{\lambda}) = f(\underline{x}) + \sum_{i=1}^{m} \lambda_i(b_i - g_i - S_i) \\ &\text{w.r.t. } x_1, x_2, \ldots x_n \\ &\qquad S_1, S_2, \ldots S_m \\ &\qquad \lambda_1, \lambda_2, \ldots \lambda_m \\ &\text{subject to } x_j \geq \ell_j \quad j + 1, 2, \ldots n. \end{aligned} \tag{26}$$

(Recall that the notation $\underline{x}$ stands for $(x_1, x_2, \ldots x_n)$, $\underline{S}$ stands for $(S_1, S_2, \ldots S_m)$, and $\underline{\lambda}$ stands for $(\lambda_1, \lambda_2, \ldots \lambda_m)$.) Combining the results from solving problems (19) and (6), the necessary conditions for the solution of problem (25) are

$$\begin{aligned} &\text{(a)} \quad \frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} \leq 0; \quad (x_j^* - \ell_j) \frac{\partial L}{\partial x_j} = 0 \\ &\qquad x_j \geq \ell_j \quad j = 1, 2, \ldots n \\ &\text{(b)} \quad \lambda_i \geq 0; \quad \lambda_i^*(b_i - g_i(\underline{x}^*)) = 0 \quad i = 1, 2, \ldots m \\ &\text{(c)} \quad g_i(x) \leq b_i \quad i = 1, 2, \ldots m. \end{aligned} \tag{27}$$

Equation (27) gives the famous Kuhn-Tucker conditions for the solution of problem (25). As in problem (9), if there is an upper limit to any activity such as $x_k \leq u_k$, the necessary condition corresponding to (27a) is

$$\frac{\partial L}{\partial x_k} = \frac{\partial f}{\partial x_k} - \sum_{i=1}^{m} \lambda_i \frac{g_i}{x_k} \geq 0, \; (x_k{}^* - u_k) \frac{\partial L}{\partial x_k} = 0. \qquad (27a)$$

$$x_k \leq u_k$$

The extension of the minimization problem (11) is similarly generalized to more variables and inequality constraints. The problem is

$$\begin{aligned} &\text{Minimize } k(y_1, y_2, \ldots y_m) \\ &\text{w.r.t. } (y_1, y_2, \ldots y_m) \qquad (28) \\ &\text{subject to } h_j(y_i, y_2, \ldots v_m) \geq c_j \quad j = 1, 2, \ldots n \\ &\qquad y_i \geq \ell_i \qquad i = 1, 2, \ldots m. \end{aligned}$$

Again follow Baumol's advice and for a minimization problem write the constraints in the form $c_j - h_j \leq 0$. Note carefully the difference in the direction of the inequalities for the minimization problem (28) as compared to that of the inequalities for the maximization problem (25). Then add a nonnegative slack variable, S_j, to each $c_j - h_j$ and form the Lagrangian using μ_j as the symbol for the jth Lagrange multiplier

$$L(\underline{y}, \underline{S}) = k(y) + \sum_{j=1}^{n} \mu_j(c_j - h_j(\underline{y}) + S_j). \qquad (29)$$

The Kuhn Tucker conditions for this minimization problem are

(a) $\frac{\partial L}{\partial y_i} = \frac{\partial k}{\partial y_i} - \sum_{j=1}^{n} \mu_j \frac{\partial h_j}{\partial y_i} \geq 0; \qquad (y_i{}^* - \ell_i) \frac{\partial L}{\partial y_i} = 0$

$y_i \geq \ell_i \qquad i = 1, 2, \ldots m.$

(b) $\mu_j \geq 0; \; \mu_j{}^*(c_j - h_j(y_8)) = 0 \qquad j = 1, 2, \ldots n$

(c) $h_j(\underline{y}) \geq c_j \qquad j = 1, 2, \ldots n.$

Footnote to Appendix A

[1] I digress to show mathematically that at the optimal values of the activities, x_1 and x_2, and the shadow price, λ, the shadow price equals the total derivative of the objective, f, with respect to b and also equals the partial derivative of the Lagrangian with respect to b. This is a special case of the "envelope theorem" and I follow the presentation of Varian (1978, p. 268).

Let $f^*(b)$ be the optimal value of $f(x_1, x_2)$, which is attained at $x_1^*(b)$, $x_2^*(b)$, $\lambda^*(b)$, and let g^* be the value of g at these values. Note that the optimal values are functions of b, the amount of resource available. The problem is given in equation (3), the Lagrangian problem in equation (4), and the necessary conditions in equation (5). By direct differentiation as follows

$$df^*(b)/db = (\partial f^*/\partial x_1)\ (dx_1^*/db) + (\partial f^*/\partial x_2)\ (dx_2^*/db) + \partial f^*/\partial b.$$

But $\partial f^*/\partial b = 0$ since f^* is not directly a function of b. From the necessary conditions of equation (5), $\partial f^*/\partial x_1 = \lambda^* \partial g^*/\partial x_1$ and $\partial f^*/\partial x_2 = \lambda^* \partial g^*/\partial x_2$. Therefore, $df^*/db = \lambda^*(\partial g^*/\partial x_1 + \partial g^*/\partial x_2)$. Since the optimal values x_1^* and x_2^* must identically satisfy the constraint so that $g(x_1^*, x_2^*) \equiv b$, differentiating this identity gives

$$(\partial g^*/\partial x_1)\ (dx_1^*/db) + (\partial g^*/\partial x_2)\ (dx_2/db) = \partial b/\partial b = 1,$$

which means that

$$df^*/db = \lambda^*.$$

Finally, observe that for all x and λ, $\partial L/\partial b = \partial(f + \lambda(b - g))/\partial b = \partial f/\partial b + \lambda - \lambda \partial g/\partial b = \lambda$, so in particular $\partial L^*/\partial b = \lambda^*$.

Appendix B: Representing the Problem

Two types of variables are used to model the dynamic decision problem. The state variables represent the condition of the system at a given time. Many of the state variables in economic and business systems are stocks such as the weight or volume of a commodity. But sometimes a state variable is a flow such as the velocity of a stream. The other types of variables are the control variables or activities that the decision maker can change directly. The objective function is to be optimized with respect to the control variables. Control variables frequently are flow variables such as the rate at which a factor is used. Sometimes, however, controls are stock variables such as the opening or closing of a gate.

i) System of Equations of Motion

For discrete time periods, the state variables in period t are written

$$\underline{K}_t = (K_{1t}, K_{2t}, \dots K_{mt}) \qquad t = 1, 2, \dots, T$$

and the control variables in period t are written

$$\underline{u}_t = (u_{1t}, u_{2t}, \dots, u_{nt}) \qquad t = 1, 2, \dots T.$$

For continuous time, the state variables at time t are written

$$\underline{K}(t) = (K_1(t), K_2(t), \dots K_m(t))$$

and the control variables at time t are written

$$\underline{u}(t) = (u_1(t), u_2(t) \dots, u_n(t)) \qquad 0 \le t \le T,$$

with the notation $\vec{\underline{u}}$ used to represent all values of $\underline{u}(t)$.

The equations of motion of the system represent the change of the state variables as functions of the level of the state variables and the control variables. For discrete periods, the change in state variable k from period t to period t+1 is written as the function

$g_{kt}(\underline{K}_t, u_t)$

$$K_{k\ t+1} - K_{kt} = g_{kt}(\underline{K}_t, \underline{u}_t) \qquad k = 1, 2, \ldots m,$$

where the subscript t on g means that g may also be a function of the time period considered. The values of the state variables must be known at some specified time. For all examples in this report, the starting values are known and are written

$$K_{k1} = K_{k0} \qquad k = 1, 2, \ldots m.$$

For continuous time, the rate of change in the state variable k is written as the total derivative of K_k with respect to time

$$\frac{dK_k(t)}{dt} = g_k(\underline{K}(t), \underline{u}(t), t) \qquad k = 1, 2, \ldots m.$$
$$0 \leq t \leq T \quad .$$

In continuous time, the initial values of the state variables at t=0 are known and these values are written

$$K_k(0) = K_{k0} \qquad k = 1, 2, \ldots m.$$

ii) The Objective Functional[1]

For each discrete time period there is a function that gives the amount added to the initial objective of the decision maker. This increment may be a function of the state variables and the control variables chosen at that time increment as well as the time of the period. This increment is written

$$V_t(\underline{K}_t, \underline{u}_t).$$

Allowing this to be a function of the period t allows for discounting form period t back to the time of the decision at the initial time. With the assumption that the effects from the different periods are additive and separable, the quantity that the decision maker seeks to optimize is

$$\sum_{t=1}^{T} V_t(\underline{K}_t, \underline{u}_t, t) \ .$$

The term separable means that the value of a variable in one period does not directly affect the value of V in another period. The optimization is with respect to the values $\underline{u}_t$, t = 1, 2, ... T.

For continuous time, there is at each moment a function for the rate at which the objective is being changed such that the objective to be optimized is the integral of this function. This function may also be directly a function of time. This fact allows discounting. The objective must be separable and the generalization of additivity means that the quantity that the decision maker seeks to optimize is the integral

$$\int_0^T V(\underline{K}(t), \underline{u}(t), t)dt.$$

The optimization is with respect to the values $\underline{u}(t)$ for all $0 \leq t \leq T$.

It is important to remember that the optimization is with respect to all of the values of the control variables over the time horizon out to time T.

iii) Constraints

As with the one-period optimization problems reviewed in Appendix A, the problem often requires inequality constraints on the control variables (activities). When there are R such constraints, write

$$h_i(\underline{u}(t), t) \geq 0 \qquad i = 1, 2, \ldots p,$$

so that in general the control constraint may depend on the period t but not on the state variables. Adding constraints on the state variables introduces more complications (see Kamien and Schwartz, pp. 215-225). Constraints on control variables often arise because the procedure divides the use of a single resource among several uses. For example, suppose there are three uses and the shares in each of the uses are $u_1 \geq 0$, $u_2 \geq 0$, $u_3 = 1 - u_1 - u_3 \geq 0$.

Constraints on the state variables at the terminal time are often required. For example, the net value of assets may have to be non-negative at the end of the planning horizon. These end point conditions can be included without much complication because the Kuhn-Tucker conditions can be applied to the shadow price at the end of time. If there are s of these terminal constraints on the state variables, with discrete time increments write

$$f_\ell(\underline{K}_{T+1}) \geq 0 \qquad \ell = 1, 2, \ldots s$$

and with continuous time write

$$f_\ell(K(T)) \geq 0 \qquad \ell = 1, 2, \ldots s.$$

In discrete time increments, the maximization problem is

$$\text{Max} \qquad \sum_{i=1}^{T} V_t(\underline{K}_t, \underline{u}_t)$$

$$\text{w.r.t.} \qquad \underline{u}_1, \underline{u}_1, \underline{u}_2, \ldots \underline{u}_T$$

subject to

$$K_{k\ t+1} - K_{kt} = \underline{g}_{kt}\ (\underline{K}_t, u_t) \qquad k = 1, 2, \ldots m$$

$$t = 1, 2, \ldots T$$

$$K_{k1} = K_{k0}$$

$$h_{it}(\underline{u}_t) \geq 0 \qquad i = 1, 2, \ldots p$$

$$f_\ell(\underline{K}_{T+1}) \geq 0 \qquad \ell = 1, 2, \ldots s.$$

In continuous time, the maximization problem is

$$\text{Max} \qquad \int_0^T V(\underline{K}(t), \underline{u}(t), t)dt$$

$$\text{w.r.t.} \quad \vec{\underline{u}}$$

subject to

$$\frac{dK_k(t)}{dt} = g_k(\underline{K}(t), \underline{u}(t), t) \qquad k = 1, 2, \ldots m$$

$$t = 1, 2, \ldots T$$

$$K_k(0) = K_{k0} \qquad k = 1, 2, \ldots m$$

$$h_i(\underline{u}(t), t) \geq 0 \qquad i = 1, 2, \ldots p$$

$$f_\ell(\underline{K}(T)) \geq 0 \qquad \ell = 1, 2, \ldots s.$$

Appendixes C and D show how to solve this problem.

Footnote to Appendix B

[1]The term functional is used because in continuous time problems there are an infinite number of time values in any finite interval and this causes the relationship between the objective and the controls to go beyond the usual definition of a function. With discrete periods, a similar situation arises when the horizon T goes to infinity.

Appendix C. Discrete Time Maximum Principle (Fixed End Time)

As shown in Appendix B, with discrete time increments the maximization problem is

$$\text{Max} \qquad \sum_{i=1}^{T} V_t(\underline{K}, \underline{u}_t)$$

$$\text{w.r.t.} \qquad \underline{u}_1, \underline{u}_2, \ldots \underline{u}_T$$

subject to

$$K_{k\ t+1} - K_{kt} = \underline{g}_{kt}(\underline{K}_t, \underline{u}_t) \qquad k = 1, 2, \ldots m$$

$$t = 1, 2, \ldots T$$

$$K_{k1} = K_{k0}$$

$$h_{it}(\underline{u}_t) \geq 0 \qquad i = 1, 2, \ldots p$$

$$f_\ell(\underline{K}_{T+1}) \geq 0 \qquad \ell = 1, 2, \ldots s.$$

To solve the discrete time optimal control problem given in Appendix B, form the Lagrangian

$$L = \sum_{t=1}^{T} V_t(\underline{K}_t, \underline{u}_t, t)$$

$$+ \sum_{k=1}^{m} \sum_{t=1}^{T} \lambda_{kt}(g_{kt} + K_{kt} - K_{k\ t+1})$$

$$+ \sum_{i=1}^{p} \sum_{t=1}^{T} \mu_{it}\, h_{it} + \sum_{\ell=1}^{s} \sigma_\ell\, f_\ell \ .$$

Here λ_{kt} is the shadow price of the change in resource k at time t and is associated with the equality constraint of the k^{th} equation of motion at time t, $g_{kt} + K_{kt} - K_{k\ t+1} = 0$. In control problems the

λ shadow prices are called costate variables or adjoint variables because they are associated with the rates of change in the state variables. The multiplier μ_{it} is the shadow price associated with the i^{th} limit on the controls at time t; and σ_{ℓ} is the shadow price of the ℓ^{th} constraint on the state variables at the end of the plan. Since there are a finite number of K's, x's, λ's, μ's, and σ's in the problem, the necessary conditions for optimization are just those developed in Appendix A. Although the state variables, K's, are not directly under the control of the optimizer, it is still necessary that the partial derivatives equal zero since the K's can be changed.

The partial derivatives with respect to the state variables are required even though only the control variables can be changed directly by the decision maker. It would be very awkward to try to include the indirect effects of the control variables through the state variables in any other way. Cannon, Cullum and Polak (1970) present a rigorous treatment of discrete control, whereas Clark (1976) just presents the equations without elaboration. The presentation in this appendix takes a middle road to give some explanation without all of the mathematical details such as in Cannon, Cullum and Polak. Sometimes it helps to use the approximation $\lambda_{kt} - \lambda_{k\ t-1} \doteq \lambda_{k\ t+1} - \lambda_{kt}$ so that $\lambda_{k\ t+1}$ can be calculated from values known from period t. The names given to each group of equations are those that have been adopted in the control literature.

(a) <u>Costate or Adjoint Conditions</u> (so called because they give the criteria for the rate of change of the costate or adjoint variables). The notation k* is used to represent the index of summation when there might be confusion using k in two different ways in one equation.

$$\frac{\partial L}{\partial K_{kt}} = \frac{\partial V_t}{\partial K_{kt}} + \sum_{k*=1}^{m} \left(\lambda_{k*t} \frac{\partial g_{k*t}}{\partial K_{k*t}}\right) + \lambda_{kt} - \lambda_{k\ t-1} = 0$$

or

$$-(\lambda_{kt} - \lambda_{k\ t-1}) = \frac{\partial}{\partial K_{kt}} \left(V_t + \sum_{k*=1}^{m} \lambda_{k*} g_{k*t}\right)$$

$$k = 1, 2, \ldots m$$

$$t = 2, 3, \ldots T \quad .$$

Since the first period value of each state variable, K_{k1}, is a given constant, there are no derivatives with respect to

K_{k1} $\quad k = 1, 2, \ldots m.$

(b) Transversality Conditions (so called because they are the criteria to be met when the system reaches the "transversal" or end time).

$$\frac{\partial L}{\partial K_{k\ T+1}} = -\lambda_{kT} + \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial K_{k\ T+1}} = 0$$

or

$$\lambda_{kT} = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial K_{k\ T+1}} . \qquad \sigma_\ell \geq 0 \quad \sigma_\ell f_\ell = 0$$

$$\ell = 1, 2, \ldots s.$$

This means that the end time value of the k^{th} costate variable is given by the sum of the worth of the marginal effects of the state variable on the terminal constraint functions, f's.

(c) Hamiltonian Conditions (so called after the mathematician Hamilton and the important use of the Hamiltonian function introduced in the next paragraph).

$$\frac{\partial L}{\partial u_{jt}} = \frac{\partial V_t}{\partial u_{jt}} + \sum_{k=1}^{m} \lambda_{kt} \frac{\partial g_{kt}}{\partial u_{jt}} \qquad \begin{array}{l} j = 1, 2, \ldots n \\ t = 1, 2, \ldots T \end{array}$$

$$+ \sum_{i=1}^{p} \mu_{it} \frac{\partial h_{it}}{\partial u_{jt}} = 0$$

with $\mu_{it} \geq 0$ and $\mu_{it} h_{it} = 0$ $\qquad i = 1, 2, \ldots p$

(Remember that this means that either $\mu_{it} = 0$ or $h_{it} = 0$)

or

$$\frac{\partial}{\partial x_{jt}} [V_t + \sum_{k=1}^{m} \lambda_{kt}\, g_{kt}] + \sum_{i=1}^{p} \mu_{it} \frac{\partial h_{it}}{\partial u_{jt}} = 0;$$

$$\mu_{it} \geq 0;\ \mu_{it} h_{it} = 0.$$

(d) Equations of Motion (so called because the original system of equations for the rates of change in state variables are regained).

$$\frac{\partial L}{\partial \lambda_{kt}} = g_{kt} + K_{kt} - K_{k\ t+1} = 0$$

or

$$K_{k\ t+1} - K_{kt} = g_{kt} \qquad \begin{array}{l} k = 1, 2, \ldots m \\ t = 1, 2, \ldots T \end{array}$$

as in the original equations describing the system dynamics.

The Hamiltonian

Notice how, in the alternate forms given above, the function $V_t + \sum_{k=1}^{m} \lambda_{kt}\, g_{kt}$ appears in the Costate Conditions and in the Hamiltonian Conditions. Notice also how this function is similar to a Lagrangian for a given time t. As an aid to notation and memory, define the Hamiltonian at time t to be

$$H_t = V_t + \sum_{k=1}^{m} \lambda_{kt} g_{kt}.$$

The necessary conditions can then be written

(a) Costate Conditions

$$-(\lambda_{kt} - \lambda_{k\ t-1}) = \frac{\partial H_t}{\partial K_{kt}} \qquad \begin{array}{l} k = 1, 2, \ldots m \\ t = 2, 3, \ldots T \end{array}$$

(b) Transversality Conditions

$$\lambda_{kT} = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial K_{k\ T+1}} \qquad \begin{array}{l} k = 1, 2, \ldots m \\ \sigma_\ell \geq 0 \quad \sigma_\ell f_\ell = 0 \\ \ell = 1, 2, \ldots s \end{array}$$

(c) Hamiltonian Conditions

$$\frac{\partial H_t}{\partial u_{jt}} + \sum_{i=1}^{p} \mu_{it} \frac{\partial h_{it}}{\partial u_{jt}} = 0 \qquad \begin{array}{l} \mu_{it} \geq 0; \qquad \mu_{it} h_{it} = 0\ , \\ j = 1, 2, \ldots n \\ t = 1, 2, \ldots T \\ i = 1, 2, \ldots p \end{array}$$

(d) Equations of Motion

$$K_{k\ t+1} - K_{kt} = \frac{\partial H_t}{\partial \lambda_{kt}} ,$$

since $\partial H/\partial \lambda_{kt} = g_{kt}$. Notice the symmetry between the costate conditions and the equations of motion. The costate conditions say that the rate at which a costate variable decreases must equal the partial derivative of H with respect to the corresponding state variable. The equations of motion say that the rate at which a state variable increases must equal the partial derivative of H with respect to the corresponding costate variable.

These same conditions are necessary for a solution to the following problem.

At each time t

Max H_t

w.r.t. $\underline{u}_t$

subject to

$$-(\lambda_{kt} - \lambda_{k\ t-1}) = \frac{\partial H_t}{\partial K_{kt}} \qquad k = 1, 2, \dots m \quad t = 2, 3, \dots T$$

$$\lambda_{kT} = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial K_{k\ T+1}} \qquad \sigma_\ell \geq 0 \quad \sigma_\ell f_\ell = 0 \quad \ell = 1, 2, \dots s$$

$$K_{k\ t+1} - K_{kt} = \frac{\partial H_t}{\partial \lambda_{kt}} \qquad t = 1, 2, \dots T$$

$$h_{it}(\underline{u}_t) \geq 0 \ . \qquad i = 1, 2, \dots p$$

The Hamiltonian equations are obtained by forming the Lagrangian for each t

$$H_t + \sum_{i=1}^{p} \mu_{it} h_{it}$$

and taking the partial derivatives with respect to the controls at t.

$$\frac{\partial}{\partial u_{jt}} (H_t + \sum_{i=1}^{p} \mu_{it} h_{it}) = 0; \qquad \mu_{it} \geq 0; \quad \mu_{it} h_{it} = 0$$

$$j = 1, 2, \ldots n$$

$$i = 1, 2, \ldots p.$$

There are two important interpretations:

> THE HAMILTONIAN FUNCTION IN PERIOD t IS THE NET "PROFIT" FOR PERIOD t ALONE, PLUS THE WORTH OF THE CHANGES IN THE STATE OF THE SYSTEM WITH THE CHANGES VALUED AT THE SHADOW PRICES λ_{kt}.
>
> THE MAXIMUM PRINCIPLE SAYS THAT THE DECISION MAKER CAN OBTAIN THE BEST RESULTS OVER THE LIFE OF THE PROJECT IF THE NEW OBJECTIVE H IS OPTIMIZED AT EACH MOMENT. THE CATCH IS THAT TO DO WHAT IS BEST AT TIME t THE DECISION MAKER MUST KNOW THE SHADOW PRICE, λ_{kt}, FOR THE CHANGE IN EVERY STATE VARIABLE.

The shadow prices are connected over the entire horizon by the Costate Conditions and the Transversality Conditions. The main difficulty is that although the values of the state variables are known at the initial period, the known boundary conditions for the costate variables are at the end time T.

Gains from Using the Hamiltonian

The problem to mazimize the Lagrangian L with which this appendix began is a problem with one objective function, max L, and nT control variables since there are n controls, u's, in each of T periods. In the equivalent Hamiltonian form there are T problems to be solved, since for each period t = 1, 2, ... T there is the problem max H_t. This may not seem like such a bargain in the discrete time problem with a finite number of periods, but with continuous time any finite interval has an infinite number of infinitesimal periods. In the continuous time model, the maximum principle will take a problem with an infinite number of control variables and turn it into an infinite number of problems each with "only" n control variables. That is a bargain because an ordinary function of time can present the solution for all values of time. As soon as students of elementary algebra find a solution for y as a function of x, they know how to solve an infinite number of problems such as: "For this value of x what is the value of y?" It is in this sense that an infinite number of problems are solved in the maximum principle.

Salvage or Bequests

Now suppose the state variables remaining in period T+1 are worth something such as for salvage or bequests. If this worth is given by the function

$$B(\underline{K}_{T+1}, T+1),$$

the objective function is then

$$\text{Max} \sum_{t=1}^{T} V_t(\underline{K}_t, \underline{u}_t, t) + B(\underline{K}_{T+1}, T+1)$$

$$\text{w.r.t. } \underline{u}_1, \underline{u}_2, \ldots \underline{u}_T$$

and the same constraints still hold. The Hamiltonian is the same and the necessary conditions for maximization are the same, except now

$$\lambda_{kT} = \frac{\partial B}{\partial K_{k\ T+1}} + \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial K_{k\ T+1}}$$

$$\sigma_\ell \geq 0 \qquad \sigma_\ell f_\ell = 0 \qquad k = 1, 2, \ldots m$$
$$\ell = 1, 2, \ldots s.$$

Thus, the marginal contribution of the k^{th} state variable to the salvage value B is added to the shadow price of state variable K_k at time T.

Appendix D. The Maximum Principle with Continuous Time (Fixed End Time)

For this problem the notation such as $\underline{K}(t) = (K_1(t), K_2(t), \ldots K_m(t))$ is used to conserve space. If there are any questions on the meaning of notation, refer to Appendix B for analogous variables. Alternatively, the reader may, for a first reading, ignore the complications of several control and state variables and read $\underline{K}(t)$ as if it referred to a single state variable. Also, there are n control variables, $\underline{u}(t)$, that are functions of t. The entire time path of all these control variables from t = 0 to t = T is denoted by $\vec{\underline{u}}$. There are m state variables, $\underline{K}(t)$, that are functions of time with known initial values, $\underline{K}(0) = \underline{K}_0$. At each time the objective is a function of the state variables, the control variables and the time, $V(\underline{K}(t), \underline{u}(t), t)$. For the objective to be additive separable at each t now means that the total objective is the integral of $V(\underline{K}, \underline{u}, t)$ over the interval $0 \leq t \leq T$. The value of the objective attained depends on the initial state, K(0), and the time path chosen for the controls, $\vec{\underline{u}}(t)$.

These values are constrained by the rate at which the state variables can change in response to the levels of the state variables and the values chosen for the control variables. The rates of change of the state variables at time t are given by the equations of motion.

$$\frac{d\underline{K}}{dt} = \underline{g}(\underline{K}, \underline{u}, \underline{t}).$$

The constraints on the state variables at the end time T are written

$$\underline{f}(\underline{K}(T)) \geq 0,$$

and the constraints on the control variables at any time t are

$$\underline{h}(\underline{u}(t), t) \geq 0.$$

The objective functional is

$$\underset{\text{w.r.t. } \vec{u}}{\text{Max}} \int_{t=0}^{T} V(\underline{K}(t), \underline{u}(t), t)dt.$$

Therefore, in continuous time this maximization problem is

$$\underset{\text{w.r.t. } \underline{\vec{u}}}{\text{Max}} \int_{t=0}^{T} V(\underline{K}(t), \underline{u}(t), t)dt$$

subject to

$$\frac{d\underline{K}}{dt} = \underline{g}(\underline{K}, \underline{u}, t)$$

$$\underline{K}(0) = \underline{K}_o$$

$$\underline{f}(\underline{K}(T)) \geq 0$$

$$\underline{h}(\underline{u}(t), t) \geq 0 \; .$$

The Hamiltonian for this problem is

$$H(\underline{K}(t), \underline{u}(t), t) = V(\underline{K}, \underline{u}, t) + \sum_{k=1}^{m} \lambda_k(t)\, g_k(t),$$

which is of the same form as developed in Appendix C for discrete time periods. The maximum principle says that the above problem will be solved if the following problem is solved for each t between 0 and T.

$$\underset{\text{w.r.t. } \underline{u}(t)}{\text{Max } H} \qquad \text{at each } 0 \leq t \leq T$$

together with

$$\frac{d\underline{\lambda}}{dt} = - \frac{\partial H}{\partial \underline{K}} \qquad \underline{\lambda}(T) = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial \underline{K}(T)}$$

$$\sigma_\ell \geq 0 \qquad \sigma_\ell f_\ell = 0$$

$$\frac{d\underline{K}}{dt} = \frac{\partial H}{\partial \underline{\lambda}} \qquad \underline{K}(0) = K_o$$

$$\underline{f}(\underline{K}(T)) \geq 0$$

$$\underline{h}(\underline{u}(t), t) \geq 0 \; .$$

To include the last two constraints, form the Lagrangian for each t

$$\mathcal{L}(\underline{K}(t),\ \underline{u}(t),\ t) = H + \sum_{i=1}^{p} \mu_i(t)h_i(\underline{u}) + \sum_{\ell=1}^{s} \sigma_\ell f_\ell$$

and solve the problem

$$\text{Max} \qquad \mathcal{L}(\underline{K}(t),\ \underline{u}(t),\ t)$$

$$\text{w.r.t. } \underline{u}(t)$$

together with

$$\frac{\partial H}{\partial \underline{K}} = -\frac{d\underline{\lambda}(t)}{dt} \qquad \underline{\lambda}(T) = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial \underline{K}(T)}$$

$$\sigma_\ell \geq 0 \qquad \sigma_\ell f_\ell = 0$$

$$\ell = 1, 2, \ldots s$$

$$\frac{\partial H}{\partial \underline{\lambda}} = \frac{d\underline{K}}{dt} \ . \qquad \underline{K}(0) = \underline{K}_o$$

With this form it is now clear that the necessary conditions to solve the original problem are, for each $0 \leq t \leq T$,

$$\frac{\partial \mathcal{L}}{\partial \underline{u}} = \frac{\partial H}{\partial \underline{u}} + \sum_{i=1}^{p} \mu_i \frac{\partial h_i}{\partial \underline{u}} = 0 \qquad \mu_i(t) \geq 0$$

$$\mu_i(t)\ h_i(\underline{u}(t)) = 0$$

$$i = 1, 2, \ldots p$$

$$\frac{d\underline{\lambda}}{dt} = -\frac{\partial H}{\partial \underline{K}} \qquad \underline{\lambda}(T) = \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial \underline{K}(T)}$$

$$\sigma_\ell \geq 0 \qquad \sigma_\ell f_\ell = 0$$

$$\ell = 1, 2, \ldots s$$

$$\frac{d\underline{K}}{dt} = \frac{\partial H}{\partial \underline{\lambda}} \qquad \underline{K}(0) = \underline{K}_o$$

$$\underline{f}(\underline{K}(T)) = 0$$

$$\underline{h}(\underline{u}(t),\ t) \geq 0 \ .$$

Often there are direct restrictions on the control variables, such as $\underline{u}(t) \geq 0$, in addition to the constraints $h \geq 0$. Two approaches might be taken. One option is to consider the constraint $\underline{u} \geq 0$ as a very simple function $h(\underline{u})$ and include these constraints in the set of constraints $\underline{h} \geq 0$. The other approach is to apply the Kuhn-Tucker conditions directly to get

$$\frac{\partial L}{\partial \underline{u}} \geq 0, \quad \frac{\partial L}{\partial u_j} u_j = 0, \quad \mu_i \geq 0, \quad \mu_i h_i = 0 \qquad \begin{array}{l} i = 1, 2, \ldots p \\ j = 1, 2, \ldots n. \end{array}$$

For many problems, some of the state variables are worth something at the end time T. Examples of this would be the "salvage value" of machinery, the market value of buildings or the utility of bequests to one's heirs. When this end worth is given by the function $B(\underline{K}(T), T)$, the objective functional is

$$\underset{\text{w.r.t. } \vec{\underline{u}}}{\text{Max}} \quad \int_{t=0}^{T} V(\underline{K}(t), \underline{u}(t), t)dt + B(\underline{K}(T), T).$$

When the problem is subject to the conditions above,

$$\frac{d\underline{K}}{dt} = \underline{g} \qquad \underline{K}(0) = \underline{K}_0$$

$$\underline{h}(\underline{u}(t), t) \geq 0$$

$$\underline{f}(\underline{K}(T) \geq 0 ,$$

then the shadow prices of the state variables at the end time must equal the marginal contribution of $\underline{K}(T)$ to the value of B plus the sum of the values of the contributions to the constraints $f \geq 0$.

$$\underline{\lambda}(T) = \frac{\partial B}{\partial \underline{K}(T)} + \sum_{\ell=1}^{s} \sigma_\ell \frac{\partial f_\ell}{\partial \underline{K}(T)} \qquad \begin{array}{l} \sigma_\ell \geq 0 \\ \sigma_\ell f_\ell \qquad \ell = 1, 2, \ldots s. \end{array}$$

For additional cases, especially cases when the end time is open to choice, see Kamien and Schwartz, pp. 147-148.

Agricultural Research Service

North Carolina State University
at Raleigh

D. F. Bateman, Director of Research