

Half-title page, prepared by publisher

Publishers' page

Full title page, prepared by publisher

Copyright page, prepared by publisher

Dedication Page
(optional)

Preface

The primary motivation for adaptive memory programming, therefore, is to group and unify all these emerging optimization techniques for enhancement of the computational capabilities that they offer to combinatorial problems which are encountered in real life in the area of production planning and control.

We confront this pitfall technically, by introducing explicit remarks about the generality of results at appropriate places; methodologically, by accumulating enough applications for every major idea to make its validity and generality stand out; and philosophically, observing that physics moves forward most of its ideas by analogies to cleverly chosen simple systems for which profound intuitions have been formed.

Special attention has also been paid to the wave-current interaction problems. Several models are evaluated by comparing the numerical results with laboratory data. It is quite clear that these higher-order modified equations are adequate for modeling the wave propagation from deep water to shallow water. However, to apply these models in the surf zone further study of breaking waves and the proper parameterization of wave breaking processes are essential.

One of most surprising findings is coastal engineering research in the last two decades is the robustness of the shallow-water equation models in calculating the wave runup in swash zone. Although the wave breaking process is usually not considered in the shallow-water equations, with a proper tuning of the numerical dissipation as well as the bottom friction, these models can predict the time history of runup heights for various types of incident waves with impressive accuracy. These models have also been extended to examine the interactions between water waves and coastal structures, which are either impermeable or protected by a layer.

The second order wave theory must be employed to include the effects of wave drift forces and springing since they are caused by the quadratic nonlinearity. To consider ringing and wave slamming, cubic nonlinearity and higher order nonlinearity must be included in the formulation.

The factors which have contributed most towards this are the growth of the uranium industry, the acceptance of solvent extraction as a process suitable for industrial use on a large scale, the development of techniques of leaching and reduction at temperatures up to 240°C at moderate pressures, and the demand for numerous less-common metals and other elements.

To model wave slamming and ringing as mentioned above and other nonlinear phenomenon, it is necessary to undertake fully nonlinear transient analyses, usually involving numerical time marching. At present many such numerical models exist. One of common difficulties faced by these models is the procedure to track the location of free surface, especially in the case of wave breaking. Different applications associated with each method, especially in wave hydrodynamics are discussed. More than one hundred references are cited in the paper.

T. M. Chan

Contents

<i>Preface</i>	vii
1. Introduction	1
1.1 The Microeconomic Problem	1
1.2 Agricultural Optimization Problems	2
1.2.1 Food and Diet Problems	2
1.2.2 Farm and Agribusiness Management	3
1.2.3 Farm Firm Development	5
1.2.4 Production Response	6
1.2.5 Interregional and Spatial Economics	6
1.2.6 Econometrics and Statistical Applications	8
1.2.7 Policy Analysis	9
1.2.8 Nonparametric Efficiency Analysis	11
1.3 An Application	11
1.3.1 A Hyperbolic Tangent Production Function	12
1.3.2 Profit Maximization	14
2. Mathematics of Optimality	17
2.1 A Review of Expansions	17
2.2 Optimality Conditions	19
2.3 Weierstrass's Theorem	19
2.4 Multivariate Optimality	20
2.4.1 First and Second Order Necessary Conditions	20
2.4.2 Imposing a General Equality Constrained Optimum	22
2.4.3 Imposing Linear Equality Constraints	25
2.5 An Applied Example	28

2.6	Linear Inequality	29
2.7	Nonlinear Constraints	34
2.7.1	Simple Constraint	34
3.	Algorithms for Optimization	37
3.1	Univariate Newton's Method	37
3.1.1	Method of Bisection	39
3.1.2	Newton's Method	40
3.2	Finding the Multivariate Maximum	40
3.2.1	A Simple Example	43
3.2.2	Prototype Algorithm	43
3.3	Quasi-Newton Methods for Optimization	44
3.3.1	Steepest Descent	45
3.3.2	Conjugate Gradient Methods	45
3.3.3	A Numerical Example	47
3.4	Linear Equality and Inequality Constraints	52
3.4.1	Linear Equality Constraints	52
3.4.2	Linear Inequality Constraints	55
3.5	Algorithms for Nonlinear Constraints	56
3.5.1	Penalty Functions	56
3.5.2	Barrier Function	57
3.5.3	Projected Augmented Lagrangian	58
4.	Computer Code	59
4.1	Unconstrained Optimization	59
4.1.1	Precision Agriculture	59
4.1.2	Estimation of the Inverse Hyperbolic Sine Transformation	59
5.	Review of Dynamic Mathematics	63
5.1	Introduction to Dynamic Optimality	63
5.2	Review of Ordinary Differential Equation	66
5.2.1	First Order Linear Differential Equation	67
5.2.2	Nonlinear Differential Equations	70
5.2.3	Second and Higher Order Linear Differential Equations	73
5.2.4	Method of Undetermined Coefficients	75
5.2.5	Laplace Transformed	76

Contents

xi

5.3	Difference Equations	80
5.3.1	First Order Linear Difference Equations	81
5.3.2	Second Order Linear Difference Equations	83
6.	Dynamic Optimality	85
6.1	Overview of Dynamic Optimization	85
6.2	Simplified Example	86
6.3	Euler Equation	90
6.3.1	A Mathematical Digression	90
6.3.2	Definitions of a Maximum	93
6.3.3	Necessary Conditions	93
6.3.4	Interpreting the Euler Equation	99
6.3.5	Second Order Conditions	109
6.3.6	Isoperimetric Problem	112
6.3.7	Free End Value Problems	114
6.3.8	Inequality Constraints	116
6.4	Optimal Control	118
6.4.1	Basic Formulation	118
6.4.2	Sufficiency of Optimal Path	123
6.4.3	Interpretations	124
6.4.4	Several Variables	126
Appendix A	Review of Linear Algebra	127
A.1	Canonical Form and Null Spaces	127
A.2	Determinant	130
A.3	Norms	132
A.4	Matrix Factorizations	134
Appendix B	MINOS 5.6 Application	137
	<i>Bibliography</i>	139

Chapter 1

Introduction

1.1 The Microeconomic Problem

The typical definition of economics is the study of the allocation of scarce resources to satisfy unlimited and competing human wants and desires. Within this basic definition, we see the elements of a mathematical programming model. First, we have an objective function, or a measure of human wants and desires which we will attempt to maximize. Second, we have constraints or limits to simply giving humans all that they could possibly want. Implied in the definition is also the concept that we have choice variables or instruments within our control to effect this maximization.

For the consumer's problem we define those elements in the mathematical formulation

$$\begin{aligned} \max_x & U(x) \\ \text{s.t. } & x'p \leq Y \end{aligned} \quad (1.1)$$

where $U(x)$ is the utility function, x is the vector of goods consumed, p is the price of the goods, and Y is the level of income. In this problem, the objective function is the consumer's utility, the constraint is income, and the choice variables are levels of different commodities consumed. How do we determine the optimum for this problem?

Similarly, the producer's problem becomes

$$\begin{aligned} \max_{x,y} & y'p - x'w \\ \text{s.t. } & y = F(x) \end{aligned} \quad (1.2)$$

where y is the level of outputs, p is the price of outputs, x is the level of inputs, w is the price of inputs, and $F(x)$ is the production function. As

in the consumer's problem, we have an objective function which is profit, constraints which is the production function, and choice variables which are the levels of inputs and outputs. Again, how do we determine the optimum in theory and in practice?

1.2 Agricultural Optimization Problems

This section briefly presents several optimization problems common to agricultural and resource economics following Moss [Moss (2002)].

1.2.1 Food and Diet Problems

The food and diet research can be characterized by two major focuses determining the least cost combination of foods to meet dietary needs (Stigler's "Cost of subsistence") and least cost feed ration studies. The basic application would involve minimizing the cost of a diet subject to some nutritional constraint

$$\begin{aligned} \min_x & c'x \\ \text{s.t. } & Ax \geq b \end{aligned} \quad (1.3)$$

where c is the vector of food (or feed) prices, x is a vector of the choice levels for each food, A is a matrix of the nutrients provided by each food, and b is a vector of minimum nutritional requirements. More advanced formulations of the diet problem have been developed in the guise of the household production model. The general form of the household production problem is

$$\begin{aligned} \max_{y,x} & U(y) \\ \text{s.t. } & y \leq F(x) \\ & p'x \leq I \end{aligned} \quad (1.4)$$

where $F(x)$ denotes the production relationship between purchased food-stuffs and consumable goods (y), p is the price vector for purchased food-stuffs, and I is the consumer's income level.

A linear form of the household production function can be derived by taking a first-order approximation of the vector-valued production function

$$A = \nabla_x F(x) = \begin{bmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{bmatrix} \quad (1.5)$$

where $\nabla_x f(x)$ denote the gradient vector (or the vector of derivatives) . The linearized household production model can then be specified as

$$\begin{aligned} \max_{y,x} & U(y) \\ \text{s.t. } & y \leq Ax \\ & p'x \leq I \end{aligned} \quad (1.6)$$

In addition to foodstuffs, x can be augmented to include labor.

1.2.2 Farm and Agribusiness Management

Initially, linear programming was used to find optimal crop mix. This work has grown into large extension farm planning efforts such as OK farms. These models tend to be either general linear or integer. For example, one formulation of the farm planning model can be written as

$$\begin{aligned} \max_x & c'x \\ \text{s.t. } & Ax \leq b \end{aligned} \quad (1.7)$$

where x is be a vector of possible crop alternatives (wheat, cotton, and oats), c is a conformable vector of net returns from each crop activity, A is a matrix of resource constraints

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0.2 & 0.3 & 0.1 \\ 25 & 100 & 10 \end{bmatrix} \rightarrow \begin{array}{l} \text{Land} \\ \text{Labor} \\ \text{Capital} \end{array} \quad (1.8)$$

and b is a vector of resource constraints. This basic formulation has been complicated in several ways. In the Midwest, crop yields are dependent on the length of the growing season. Further, this growing season is determined by the number of field days. However, the number of field days is stochastic because they depend on the weather. One way frequently used to control the effect of the weather is to increase the size of the equipment. Extending the machinery section and integrating a machinery and labor compliment lead to B-10 in Indiana that is a machinery-planning model.

Another extension of the mathematical programming approach has been the incorporation of risk into the farm-planning model (see Moss's Chapter 6 [Moss (2010)]). One way that risk may enter the farm management model is by complicating the objective function

$$\begin{aligned} \max_x & E[U(x)] \\ \text{s.t. } & Ax \leq b \end{aligned} \quad (1.9)$$

where $E[\cdot]$ is the expected value operator, $U(\cdot)$ is the utility function, A is the resource coefficients, b is the vector of resource constraints, and x is the level of each activity. Freund [Freund (1953)] shows that given that preferences are negative exponential and returns are normally distributed, the expected utility function becomes

$$\left. \begin{aligned} U(x) &= -\exp(-\rho x) \\ x &\sim N(\mu, \Omega) \end{aligned} \right\} \Rightarrow E[U(x)] = \mu'x - \frac{\rho}{2}x'\Omega x. \quad (1.10)$$

Therefore, the maximization problem becomes a nonlinear optimization problem

$$\begin{aligned} \max_x & \mu'x - \frac{\rho}{2}x'\Omega x \\ \text{s.t. } & Ax \leq b \end{aligned} \quad (1.11)$$

However, given that few closed form conjugates exist, technologies have evolved to allow direct optimization of more generalized problems

$$U(y) = \frac{y^\lambda}{\lambda} \quad (1.12)$$

$$y \sim N(\mu'x, x'\Omega x)$$

These formulations typically do not possess closed form solutions and have to be solved using numerical procedures such as numerical quadrature or Gaussian quadrature.

Agribusiness firms benefit from optimization techniques in several ways: Inventory Management, Transshipment Models, and Least Cost Feeds. Other applications which emphasize the dynamic nature of decision making include machinery replacement models which are based on Bellman's principle of optimality.

1.2.3 Farm Firm Development

The typical farm firm development model is primarily interest in firm growth. Multiperiod programming allowed researchers to examine questions related to investment, capital accumulation, and debt choice. Swanson [Swanson (1956)], and Loftsgard and Heady [Loftsgard and Heady (1959)] focused on the managerial or production response problem. Later Irwin and Baker [Irwin and Baker (1962)], and Martin and Plaxico [Martin and Plaxico (1967)] focused on the financial aspects of growth. Moss [Moss (1984)] focused on the effect of tax codes on firm growth while Moss [Moss *et al.* (1987)] focused on the effect of exogenous factors such as macroeconomic innovations on firm growth. Early studies in this area were deterministic, obviously excluding an important aspect of firm growth. Johnson, Tefertiller, and Moore [Johnson *et al.* (1967)] opened the door by allowing crop yields to be stochastic. Moss [Moss *et al.* (1987)] drew from other techniques to approximate stochastic optimal control. A significant problem with these studies have typically been the size of the matrix.

A simple linear formulation of the firm growth model can be specified as

$$\begin{aligned} & \max \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \\ \text{s.t.} \quad & \begin{bmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ T_{12} & A_{22} & 0 & \cdots & 0 \\ 0 & T_{23} & A_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \leq \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} \end{aligned} \quad (1.13)$$

Focusing on the first two constraints

$$\begin{aligned} A_{11}x_1 &\leq b_1 \\ T_{12}x_1 + A_{22}x_2 &\leq b_2 \Rightarrow A_{22}x_2 \leq b_2 - T_{12}x_1 \end{aligned} \quad (1.14)$$

Again generalizing the model, a decision in year 1 may have multiple possible outcomes

$$\begin{aligned}
& \max \begin{bmatrix} c_{11} & p_1 c_{12} & p_2 c_{22} & p_3 c_{32} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{22} \\ x_{32} \end{bmatrix} \\
& s.t. \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ T_{12} & A_{12} & 0 & 0 \\ T_{22} & 0 & A_{22} & 0 \\ T_{32} & 0 & 0 & A_{32} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{22} \\ x_{32} \end{bmatrix} \leq \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}
\end{aligned} \tag{1.15}$$

p_1 denotes the probability of event 1 occurring, p_2 is the probability of event 2 occurring, and p_3 is the probability of event 3 occurring. If event 1 occurs the profit vector c_{12} and T_{12} resources transfer to period 2

$$\begin{aligned}
A_{12}x_{12} &\leq b_{12} + T_{12}x_{11} \\
A_{22}x_{22} &\leq b_{22} + T_{22}x_{11} \\
A_{32}x_{32} &\leq b_{32} + T_{32}x_{11}
\end{aligned} \tag{1.16}$$

1.2.4 Production Response

Production response models have been used to study the impact of some policy or external shock to the sector. From the firm level, the effect of changing fertilizer prices, labor availability, or support prices on firm outputs, profits, and input demands can be mapped out, much like the duality approach to production. The firm level effects are then aggregated to the sector level. An example of this type of model is Center for Agricultural and Rural Develop (CARD).

1.2.5 Interregional and Spatial Economics

A large section of the literature, but I only want to discuss two types: **Arbitrage or regional supply and demand models:** If demand were distributed identically across all areas of the country, but there was a primary supply region (for example the orange juice market). What would the equilibrium look like? What is the impact of profit on a change in transportation costs? What happens if Brazil enters the market? And **Plant location models:** In the past most heavy industry was located in the north, particularly in the area around the great lakes. As the population shifted south more industry shifted south. Why? Current topics, the technology is being developed to condense most of the water out of milk,

reducing its shipping cost. Will this cause the dairies to move to Wisconsin? If you were John Deere how do you determine where to locate your warehouses?

Leibnitz's Rule

$$V(r) = \int_{A(r)}^{B(r)} f(x, r) dx$$

$$V'(r) = \frac{\partial V(r)}{\partial r} = f(B(r), r) \frac{\partial B(r)}{\partial r} - f(A(r), r) \frac{\partial A(r)}{\partial r} + \int_{A(r)}^{B(r)} \frac{\partial f(x, r)}{\partial r} dx \quad (1.17)$$

The competitive market equilibrium maximizes the total consumer and producer surplus. Mathematically, in a competitive market equilibrium

$$\max_x \int_0^x p^d(z) dz - \int_0^x p^s(z) dz \quad (1.18)$$

where $p^d(z)$ is the consumer's inverse demand curve and $p^s(z)$ is the producer's supply curve (i.e., the area under the producer's supply curve represents the total variable cost of production). Differentiating the sum of consumer surplus and producer surplus yields

$$p^d(z) - p^s(z) = 0 \quad (1.19)$$

Extending the problem to include two demand curves and one producer

$$\begin{aligned} \max_{x_T, x_1, x_2} \quad & \int_0^{x_1} p_1^d(z) dz + \int_0^{x_2} p_2^d(z) dz - \int_0^{x_T} p^s(z) dz - tx_1 \\ \text{s.t.} \quad & x_1 + x_2 \leq x_T \end{aligned} \quad (1.20)$$

$$\Rightarrow \begin{cases} \frac{\partial S}{\partial x_1} = p_1^d(x_1) - p^s(x_1 + x_2) = 0 \\ \frac{\partial S}{\partial x_2} = p_2^d(x_2) - p^s(x_1 + x_2) - t = 0 \end{cases}$$

where $p_1^d(z)$ is the inverse demand curve for the first consumer and $p_2^d(z)$ is the inverse demand curve for the second consumer.

1.2.6 Econometrics and Statistical Applications

Historically, econometrics relied on closed form solutions made possible by linear models of normally distributed random variables. More complicated models that introduce factors such as concavity constraints and nonnormality do not imply closed form solutions. For example, consider concavity constrained cost functions as estimated by Featherstone and Moss [Featherstone and Moss (1994)]. Basic cost function formulation

$$\left. \begin{array}{l} \min_x w'x \\ \text{s.t. } F(x, y) = 0 \end{array} \right\} \Rightarrow C(w, y) \quad (1.21)$$

$$C(w, y) = \alpha_0 + \alpha'w + \frac{1}{2}w'Aw + \beta'y + \frac{1}{2}y'By + w'\Gamma y$$

In this formulation, both the A and B matrices are symmetric by Young's theorem. In addition, if the optimization conditions are met (as we will describe in this course) the cost function is concave in input prices, implying that the A matrix is negative semi-definite. Next, we estimate this cost function using concentrated maximum likelihood. Specifically, the residual vector based on any set of parameter estimates is

$$e_i(\theta) = \begin{bmatrix} c_i - \left(\alpha_0 + \alpha'w_i + \frac{1}{2}w_i'Aw_i + \beta'y_i + \frac{1}{2}y_i'By_i + w_i'\Gamma y_i \right) \\ x_{1i} - (\alpha_1 + A_{11}w_{1i} + A_{12}w_{2i} + A_{13}w_{3i} + \Gamma_{11}y_{1i} + \Gamma_{21}y_{2i} + \Gamma_{31}y_{3i}) \\ x_{2i} - (\alpha_2 + A_{21}w_{1i} + A_{22}w_{2i} + A_{23}w_{3i} + \Gamma_{12}y_{1i} + \Gamma_{22}y_{2i} + \Gamma_{32}y_{3i}) \\ x_{3i} - (\alpha_3 + A_{31}w_{1i} + A_{32}w_{2i} + A_{33}w_{3i} + \Gamma_{13}y_{1i} + \Gamma_{23}y_{2i} + \Gamma_{33}y_{3i}) \end{bmatrix} \quad (1.22)$$

these parameters are estimated by maximizing

$$\begin{array}{ll} \max_{\theta} & -\frac{T}{2} \ln |\Omega(\theta)| \\ \text{s.t. } & \Omega(\theta) = \frac{1}{T} \sum_{i=1}^T e_i(\theta)' e_i(\theta) \end{array} \quad (1.23)$$

While this formulation is complex, it can be estimated using iterative generalized least squares without resorting to complex mathematical programming algorithms. However, one approach to estimating a concavity constrained cost function is to estimate the Cholesky decomposition of the A matrix. Specifically,

$$\begin{aligned}
 A &= \begin{bmatrix} \alpha_{11} & 0 & 0 \\ \alpha_{12} & \alpha_{22} & 0 \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ 0 & \alpha_{22} & \alpha_{23} \\ 0 & 0 & \alpha_{33} \end{bmatrix} \\
 &= \begin{bmatrix} \alpha_{11}\alpha_{11} & \alpha_{11}\alpha_{12} & \alpha_{11}\alpha_{13} \\ \alpha_{11}\alpha_{12} & \alpha_{12}\alpha_{12} + \alpha_{22}\alpha_{22} & \alpha_{12}\alpha_{13} + \alpha_{22}\alpha_{23} \\ \alpha_{11}\alpha_{13} & \alpha_{12}\alpha_{13} + \alpha_{22}\alpha_{23} & \alpha_{13}\alpha_{13} + \alpha_{23}\alpha_{23} + \alpha_{33}\alpha_{33} \end{bmatrix}
 \end{aligned} \tag{1.24}$$

This specification is nonlinear enough to require the use of mathematical programming algorithms.

1.2.7 Policy Analysis

Depending on the scope of the policy analysis, mathematical programming analysis may imply either partial or general equilibrium formulations. Partial equilibrium analysis analyzes the impact of policy changes (such as changes to price floors or export enhancement) on the market-clearing price within a single market. Recent studies have analyzed the impact of additional identity preservation costs imposed on marketing channels by the release of genetically modified organisms. General equilibrium analysis examines the impact of policy changes on the system of prices in the economy. The only problem is that we know that other prices change as we remove the price floor.

As depicted in Figure 1.1, the demand for good 2 shifts outward if good 1 and good 2 are substitutes (for the consumer the effect of the removal of the price floor is a price increase.) On the supply side, the decreased price of good 1 leads to an outward shift in supply for good 2 given that the production of good 1 and good 2 compete for limited resources. This interaction is complicated as the first market reacts. Specifically, the increased price of good 2 leads to a leftward shift in the supply curve in the market for good 1 and an upward shift in the demand for good 1 as depicted in Figure 1.2. These interactions are further complicated by changes in the factor market. Specifically, if the industry represented a significant demand for the variable factor, the initial elimination of the price floor implies a decline in the value of the input. Hence, the overall demand for the input shifts to the left.

Modeling the interaction between the two markets involves moving from a partial equilibrium analysis to a general equilibrium analysis. Early work

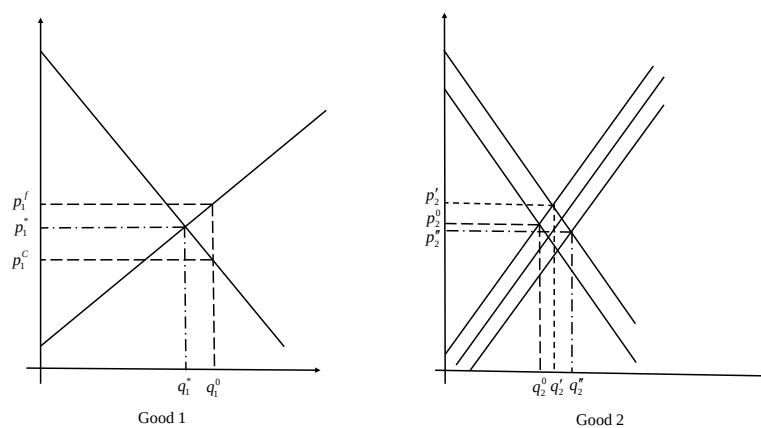


Fig. 1.1 General Equilibrium

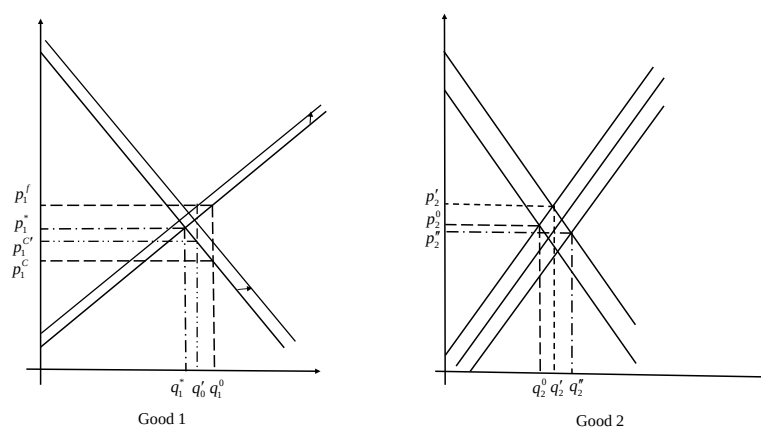


Fig. 1.2 General Equilibrium - Shift 1

on general equilibrium analysis involves the concept of a Walrasian equilibrium. The primary idea of the Walrasian equilibrium was the concept that some price vector could be found for any endowment that equated the supply and demand, or resulted in zero excess demand:

$$\begin{aligned}
\xi_i(p) &= D(p, W) - S(p, w) \\
p_i(\xi_i(p) - W_i) &= 0 \\
\xi_i(p) - W_i &\leq 0
\end{aligned} \tag{1.25}$$

$\xi_i(p)$ is the excess demand for good i , this excess demand is a function of the price vector. The demand is determined in part by the initial endowment of goods W . $p_i(\xi_i(p) - W_i)$ is the complementary slackness conditions. This condition implies that either the price of the i^{th} good is zero, or its excess demand is zero.

1.2.8 Nonparametric Efficiency Analysis

A final agricultural specification that has been growing in popularity is the nonparametric analysis of production efficiency. Economic efficiency of farms and agribusinesses has been analyzed by first estimating a parametric cost or profit function as developed in [Featherstone and Moss (1994)]. However, the efficiency results in these studies are conditioned on the choice of structural form used to estimate the parametric production function as well as the distributional assumptions used in estimation. An alternative approach is to allow the most efficient firms to form an efficient technological envelope. Assuming that a firm produces a vector of m output from n inputs, a vector of outputs (y^*) could be produced using some combination of outputs from firms in the sample.

Mathematical Specification

$$\begin{aligned}
\min_z c^* &= \sum_{i=1}^n c_i z_i \\
\text{s.t. } \sum_{i=1}^n y_{ij} z_i &\geq y_j^*, j = 1, \dots, m \\
\sum_{i=1}^n z_i &= 1
\end{aligned} \tag{1.26}$$

1.3 An Application: The Case of Precision Agriculture

One of the current rages in commercial agriculture is the evolution of precision agriculture. Several equipment companies are currently producing equipment such as combines, fertilizer spreaders, and sprayers to either provide input or respond to global positioning systems. The notion is that

instead of treating fields as 100 acre units, new technologies can increase productivity by dividing the field into 30 foot by 30 foot units. Fertilizer and pesticide applications can then be tailored to these small tracts.

The whole philosophy of precision agriculture would appear sound. Specifically, the use of information to equate the marginal cost and revenue for each productive unit would appear to increase economic efficiency. However, as is typically the case, the information is not costless. The new measuring devices for combines may cost several thousand dollars per machine. In addition, the farmer faces a substantial commitment in time and must acquire additional human capital. Thus, some study of the marginal costs and potential returns to this technology would appear appropriate.

This section examines the potential payoff of precision agriculture using a stylized primal approach. Specifically, I assume that each field is actually a combination of two fields with different primal production functions. The gain to precision agriculture is then the profit from identifying each tract of land. This gain must be compared against the cost of the new equipment. To simplify the study, I assume a single fertilizer input.

1.3.1 A Hyperbolic Tangent Production Function

The primal approach requires the assumption of a production function. In this study, we have decided to use the hyperbolic tangent function. Most production functions have at least one draw back. The traditional Cobb-Douglas function possesses a constant concavity ruling out increasing yields at an increasing rate. However, the hyperbolic tangent production function also has a problem in that it does not possess a stage III (see Moss [Moss (2014)] for a discussion of the stages of production).

Graphically, the hyperbolic tangent function in its original form is presented in Figure 1.3. Thus, we must modify it to make a legitimate production function. In this study, we define corn production as

$$f(x, \alpha_0, \alpha_1, \alpha_{max}) = \frac{\alpha_{max}}{2} [1 + \tanh(\alpha_0 + \alpha_1 x)] \quad (1.27)$$

where x represents the level of fertilizer applied while the parameters α_0 , α_1 , and α_{max} allow us to control the shape of the function. The α_0 and α_1 parameters allow us to change the location of the inflection point and the relative steepness of the production function at every point while α_{max} sets the maximum yield. Letting $\alpha_0 = -1.25$, $\alpha_1 = 0.05$, and $\alpha_{max} = 150$ yields the stylistic production function in Figure 1.4.

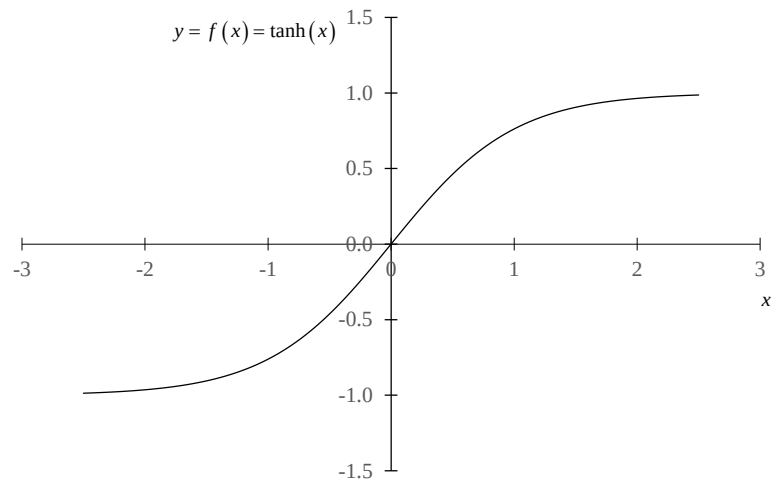


Fig. 1.3 Hyperbolic Tangent Function

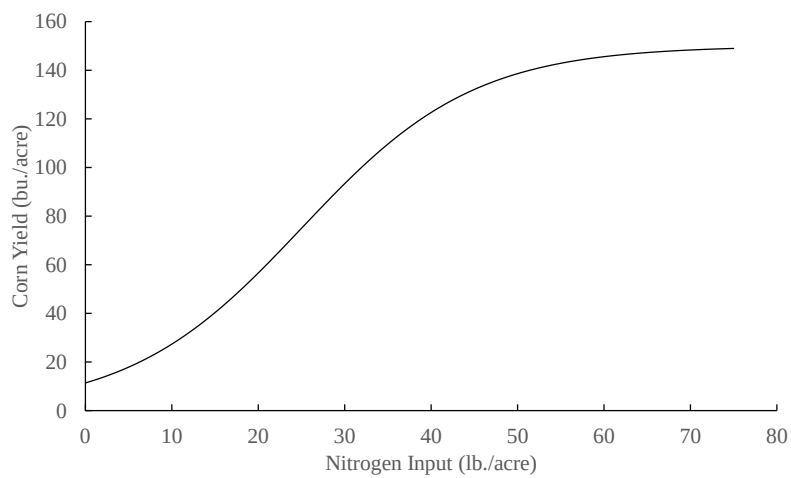


Fig. 1.4 Hyperbolic Tangent Production Function for Corn

Given this basic production function, we have several possibilities for varying production within the same field. For example, the soil may be of the same type, but past production practices may have changed the carryover level of nitrogen in the soil. In this case, it would be possible

to envision several parallel production surfaces as depicted in Figure 1.5. This figure depicts the difference between $\alpha_0 = -1.25$ and $\alpha_0 = -1.50$. The production function with $\alpha_0 = -1.50$ would represent those pieces of ground with a lower nitrogen carryover.

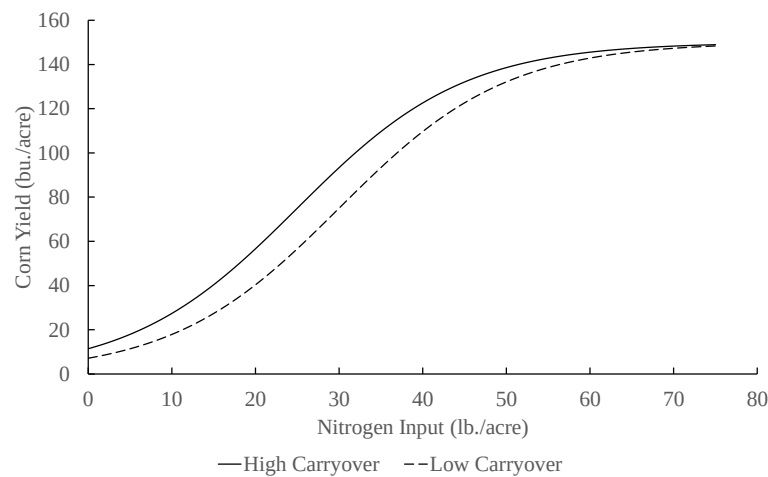


Fig. 1.5 Production Function with Different Nitrogen Carryover

An alternative is the case where the soil types are fundamentally different. For example, a field may have sand and sandy loam. In this scenario, the productive capacity of the land as well as the relative nitrogen carryover may be altered as depicted in Figure 1.6. In this scenario, we have changed from $\alpha_0 = -1.25$ to $\alpha_0 = -1.50$ and set α_{max} to 125 for the poor soil and 175 for the good soil. Intuitively, this keeps the same average maximum yield at 150 bushels per acre. The relationship between the average production functions and the production function for poor and good soils is depicted in Figure 1.7.

1.3.2 Profit Maximization

Given these production relationships, we turn to the question of: What is precision agriculture worth? We will accomplish this analysis in stages. The concept of a technology which separates a large field into several sub-fields automatically raises the question of accuracy. The first analysis will assume that the new technology separates the fields perfectly. Afterwards,

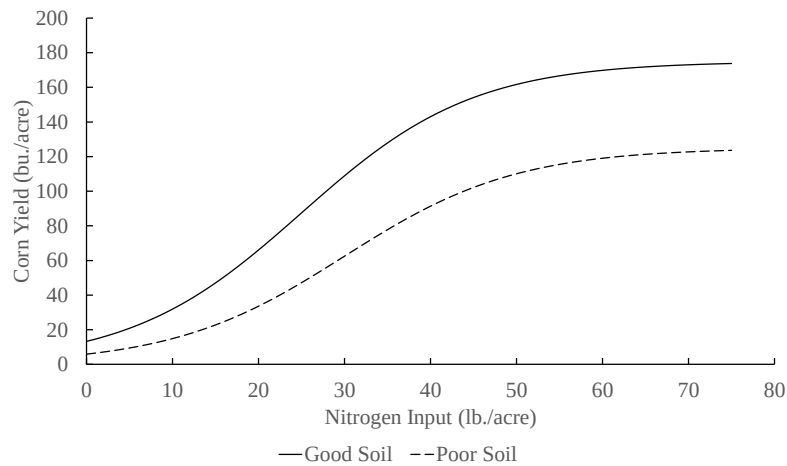


Fig. 1.6 Production Function with Different Soil Qualities

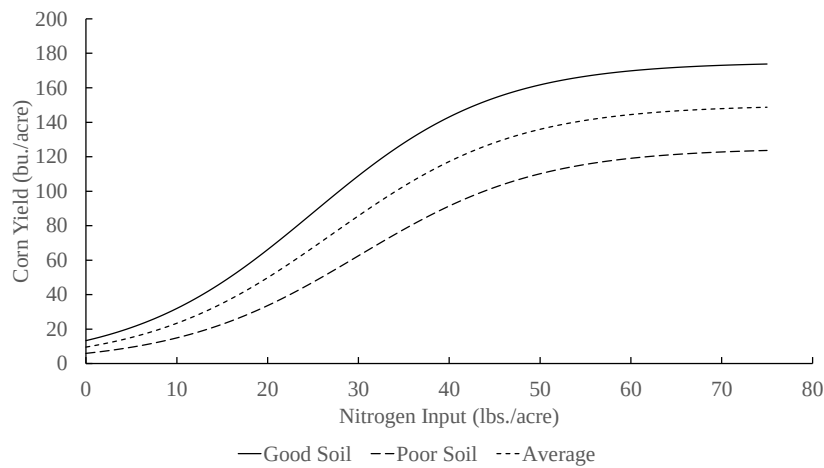


Fig. 1.7 Average Production Function for Good and Poor Soils

we will build on those results by allowing for slippage so that the high carryover subfield is treated as a low carryover subfield. This scenario may be visualized as either a inaccuracy of the technology or the limit of the unit of measure (i.e. 30 foot by 30 foot instead of 10 foot by 10 foot).

Perfect Seperation

As a first step, we begin with the assumption that both the precision technology and the associated measurement are perfect. Assuming that δ is the percent of the land of the higher carryover type, the perfectly separated optimum is derived from the profit maximization problem

$$\max_{x_1, x_2} \pi(\delta) = 3.00 [\delta f(x_1, -1.25, .05, 150) + (1 - \delta) f(x_2, -1.50, .05, 150)] - .50(\delta x_1 + (1 - \delta) x_2) \quad (1.28)$$

where we have assumed that the corn price is \$3.00/bushel with a fertilizer price of \$.50/pound. The numeric optimum for this optimization problem indicates that $x_1^* = 69.71$ pounds of fertilizer while $x_2^* = 74.77$ pounds. The profit per acre for this allocation is \$408.81.

No Separation

The profit without separation can be derived from the maximization problem

$$\max_x \pi(\delta) = 3.00 [\delta f(x, -1.25, .05, 150) + (1 - \delta) f(x, -1.50, .05, 150)] - 0.50x \quad (1.29)$$

where x is the level of fertilizer applied to both the high and low carryover parcels. The optimum level of fertilizer without separation can be derived as 72.57 pounds per acre. As we would expect, this level of fertilizer is about half way between the separated levels. The optimal level of profit without separation is \$408.66/acre. Thus, the net return to precision agriculture would be \$.15/acre or, assuming that the farmer farms two sections, \$193.63/year. The amortized value of this gain over 10 years at a discount rate of 8% is \$1299.28.

Graphically, the potential gain to precision agriculture is a function of the distribution between the two types of land. Figure 5 depicts the amortized value of precision farming as δ increases. In general, we see that the value of precision agriculture increases as δ approaches .5 and then declines. In addition, figure 5 also depicts the effect of a reduced dispersion between the production functions. Specifically, if we compare α_0 equal to -1.30 or -1.50 as opposed to -1.25 or -1.50, we see that the relative gain declines.

Chapter 2

Review of the Mathematics of Optimality

2.1 A Review of Expansions

First a review of the general notions of the Taylor expansion. The general notion of the Taylor expansion is the value of a function at a point (z) can be derived based on the value of that function at another point (a), the derivatives of the function evaluated at a , and the difference between a and z . Specifically,

$$f(z) = f(a) + f'(a)(z-a) + \frac{f''(a)}{2}(z-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(z-a)^n + \dots \quad (2.1)$$

However, the most powerful statement about the Taylor series expansion is that we know something about the residual or the approximation error if we truncate the expansion to a lower order

$$\begin{aligned} f(z) = f(a) + f'(a)(z-a) + \frac{f''(a)}{2}(z-a)^2 + \\ \dots + \frac{f^{(n)}(a)}{n!}(z-a)^n + R_{n+1}(z-z^*)^{n+1} \end{aligned} \quad (2.2)$$

$$\text{where } R_{n+1} = \frac{f^{(n+1)}(a)}{(n+1)!} \text{ for some } z^* \in [z, a]$$

Equation 2.1 states that the infinite order expansion is an equivalent function to the original function while Equation 2.2 depicts the approximation over the first n derivatives. This approximation has an error, but the error is bounded based on the $n+1$ derivative.

The first notion is reasonable and somewhat intuitive in the case of the quadratic function

$$f(x) = x^2 + ax + c \quad (2.3)$$

Using the Taylor apparatus, we approximate the value of the function at x_1 given the value of the function at x_0

$$\begin{aligned} f(x_1) &= f(x_0) + (2x_0 + a)(x_1 - x_0) + \frac{2}{2}(x_1 - x_0)^2 \\ &= \{x_0^2 + ax_0 + c\} + \{2x_0x_1 + ax_1 - 2x_0^2 - ax_0\} + \{x_1^2 - 2x_1x_0 + x_0^2\} \\ &= x_1^2 + ax_1 + c \end{aligned} \quad (2.4)$$

so in the case of the quadratic function, the second order Taylor series expansion is exact. In order to induce an error term, consider the second order expansion of a cubic function

$$\begin{aligned} g(x) &= x^3 + ax^2 + bx + c \\ g(x_1) &\approx g(x_0) + (3x_0^2 + 2ax_0 + b)(x_1 - x_0) + \frac{(6x_0 + 2a)}{2}(x_1 - x_0)^2 \end{aligned} \quad (2.5)$$

Substituting the $g(x_0)$ into the approximation yields

$$g(x_1) \approx \tilde{g}(x_1) = (3x_0 + a)x_1^2 - (3x_0^2 + b)x_1 + x_0^3 + c. \quad (2.6)$$

The approximation error can then be written as

$$\epsilon(x_1, x_0) = g(x_1) - \tilde{g}(x_1) = (x_1 - x_0)^3. \quad (2.7)$$

Hence, the approximation error for the Taylor series expansion is bounded in the neighborhood around the point of approximation.

Other useful expansions include the Fourier expansion which depicts a function as the sum of sinusoids:

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos(w_n t) + b_n \sin(w_n t)] \quad (2.8)$$

where w_i is a specific frequency. The notion is that a sufficient number of sinusoids can be added together to sufficiently proxy any general real function.

2.2 Optimality Conditions

The real workhorse in our discussion of optimality is the Taylor series expansion. First, some terminology. Let us begin by assuming that there exists a univariate function $f(x)$ defined over a given range $x \in [x_0, x_1]$. x^* is defined as a strong local minimum if $f(x^*) < f(x)$ for all $x \in N(x^*, \delta)$, $x \neq x^*$. x^* is a weak local minimum if $f(x^*) \leq f(x)$ for all $x \in N(x^*, \delta)$, $x \neq x^*$. However, local optimals are not sufficient, i.e. additional assumptions are required to guarantee that local optimal are also global.

Next, we derive the conditions for optimality from Taylor series. First, note that the value of the function $f(x)$ can be derived for any point $x \in N(x^*, \delta)$ using the Taylor series expansion of that function from the point x^* ($f(x^*)$)

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2 + R_3. \quad (2.9)$$

Assuming that $R_3 \rightarrow 0$ as $\delta \rightarrow 0$ and rearranging we have

$$f(x) - f(x^*) = f'(x^*)(x - x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2. \quad (2.10)$$

If $f(x^*) < f(x)$ for all $x \in N(x^*)$ then $f(x) - f(x^*) > 0$ for $x \in N(x^*)$. Thus, if x^* is a strong local minimum then

$$f'(x^*)(x - x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2 > 0. \quad (2.11)$$

Given that $x \in N(x^*)$ is a two sided concept $(x - x^*) \geq 0$. Thus, as $x \rightarrow x^*$ from the left $f'(x) \leq 0$. Similarly, if $x \rightarrow x^*$ from the right $f'(x) \geq 0$. If the first derivative is smooth, these conditions imply that $f'(x^*) = 0$ in order for the derivative to switch from negative to positive at x^* . In addition, if the first derivative is smooth so that the derivative switches from negative to positive, $f''(x^*) > 0$ for $x \in N(x^*)$.

2.3 Weierstrass's Theorem

The Weierstrass theorem guarantees that a continuous function has a maximum and a minimum on a closed and bounded interval. The notion of a continuous function is that there exists a function defined on an interval A such that

$$f : A \rightarrow \mathbb{R}^1 \text{ where } A = [x_0, x_1] \text{ and } \lim_{x \rightarrow x^*-} f(x) = \lim_{x \rightarrow x^*+} f(x) = f(x^*) \quad (2.12)$$

for all $x^* \in A$. Thus, there are no holes in the functional mapping.

Given this general structure we know that the maximum or minimum must either be at a local maximum or minimum, or a boundary point. The notion that the function is continuous presents the possibility of discontinuous derivatives. Thus, we are left with two general scenarios. A continuous function on a closed and bounded interval does not possess a local maximum or minimum. In this scenario either $f(x^*) - f(x) < 0$ for all $x^* > x$ in the interval in which case the function is constantly decreasing over the entire interval. In which case, the function reaches its maximum at the lower bound of the interval and its minimum at the upper bound of the interval. Alternatively, $f(x^*) - f(x) > 0$ for all $x^* > x$ in the interval in which case the function is constantly increasing over the entire interval. Hence, the function reaches its maximum at the upper bound of the interval and its minimum at the lower bound of the interval. In the second scenario, the function possesses local maxima or minima in the interval. Under this scenario, the global maximum or minimum may occur at either one of those local optima or at the boundaries. Note that we only guarantee the existence of the maximum or minimum values, not the uniqueness of x associated with those values.

Regularity Conditions The smoothness of the calculus then may allow a unique characterization. If $f(x) \in C^1$ (i.e., the first derivative of the function is continuous throughout the domain) then the relative minimum or maximum only occurs where $f'(x) = 0$. However, again, this does not guarantee a unique solution for x . Further, $f(x) \in C^2$ (i.e., the second derivative of the function is continuous throughout the domain) also does not guarantee a unique solution. However, if $f(x) \in C^2$ and $f''(x) < 0$ over the entire interval for a maximum or $f''(x) > 0$ over the entire interval for a minimum then the optima will be unique.

2.4 Introduction to Multivariate Unconstrained and Constrained Optimality

2.4.1 First and Second Order Necessary Conditions

As a first step, let us develop the first and second order necessary conditions for the multivariate function. Like the univariate case, the development of

the multivariate optimum is based on the Taylor series expansion. The multivariate form of the Taylor series expansion is typically restricted to the first two moments by convenience, as

$$f(x) = f(x^*) + \nabla_x f(x^*) dx + \frac{1}{2} dx' \nabla_{xx}^2 f(x^*) dx. \quad (2.13)$$

Again, subtracting $f(x^*)$ from each side we are left with

$$f(x) - f(x^*) = \nabla_x f(x^*) dx + \frac{1}{2} dx' \nabla_{xx}^2 f(x^*) dx \leq 0 \quad (2.14)$$

for a maximum at x^* . Thus, the first order necessary condition for the unconstrained maximum is that $\|\nabla_x f(x^*)\| \rightarrow 0$ (i.e., the norm of the gradient matrix approaches zero – basically a multivariate form of notion that the first derivative equals zero).¹ The second order necessary condition for the unconstrained maximum is that $\nabla_{xx}^2 f(x^*)$ is negative semi-definite (or $x' \nabla_{xx}^2 f(x^*) x \leq 0$) for all x in the neighborhood around x^* . The later condition is if all of the eigenvalues of the Hessian matrix are less than or equal to zero.

As an example, consider the optimization problem

$$\begin{aligned} \max_{y_1, y_2, x_1, x_2, x_3} \quad & 3.75 \exp(y_1) + 4.65 \exp(y_2) - 1.50x_1 - 2.00x_2 - 1.25x_3 \\ \text{s.t.} \quad & y_1 = f_1(x_1, x_2, x_3) = 0.15 \ln(x_1) + 0.25 \ln(x_2) + 0.55 \ln(x_3) \\ & y_2 = f_2(x_1, x_2, x_3) = 0.30 \ln(x_1) + 0.35 \ln(x_2) + 0.30 \ln(x_3) \end{aligned} \quad (2.15)$$

which is similar to maximizing the product of related Cobb-Douglas production functions. Substituting each of the constraints into the objective function yields

$$\begin{aligned} \max_{x_1, x_2, x_3} \quad & \pi = 3.75x_1^{0.15}x_2^{0.25}x_3^{0.55} + 4.65x_1^{0.30}x_2^{0.35}x_3^{0.30} \\ & - 1.50x_1 - 2.00x_2 - 1.25x_3 \end{aligned} \quad (2.16)$$

Consider the point

¹The norm of a vector can be defined in general as

$$\|x_i\|_p = \sum_{i=1}^k |x_i^p|^{1/p}$$

where $p \geq 0$. If $p = 2$, the norm is a traditional distance

$$\{x_1, x_2, x_3\} = \{8.7357, 8.8229, 20.5546\} \quad (2.17)$$

the gradient vector of Equation 2.16 becomes

$$\nabla_x \pi = \begin{bmatrix} -1.50 + \frac{0.8250x_2^{0.35}x_3^{0.30}}{x_1^{0.70}} + \frac{0.3750x_2^{0.25}x_3^{0.55}}{x_1^{0.85}} \\ -2.00 + \frac{0.9625x_1^{0.30}x_3^{0.30}}{x_2^{0.65}} + \frac{0.6250x_1^{0.15}x_3^{0.55}}{x_2^{0.75}} \\ -1.25 + \frac{0.8250x_1^{0.30}x_2^{0.35}}{x_3^{0.70}} + \frac{0.13750x_1^{0.15}x_2^{0.25}}{x_3^{0.45}} \end{bmatrix} = \begin{bmatrix} -1.6373 \times 10^{-9} \\ 4.3609 \times 10^{-9} \\ -5.97424 \times 10^{-10} \end{bmatrix}. \quad (2.18)$$

The 2-norm ($\|\nabla_x \pi\|_2$) is 4.64554×10^{-9} which is close to zero. Numerically, the Hessian matrix at this point is

$$\nabla_{xx}^2 \pi = \begin{bmatrix} -0.1294 & 0.0534 & 0.0285 \\ 0.0534 & -0.1574 & 0.0400 \\ 0.0285 & 0.0400 & -0.0323 \end{bmatrix} \quad (2.19)$$

which is negative definite (e.g., the maximum eigenvalue of the matrix in Equation 2.19 is -0.0048).

2.4.2 Imposing a General Equality Constrained Optimum

The first stage in our discussion will be a fairly general description of the problem at hand. We are now interested in developing the optimality conditions for the problem

$$\begin{aligned} \max_x & f(x) \\ \text{s.t.} & G(x) = b \end{aligned} \quad (2.20)$$

where $f(x)$ is a scalar valued function mapping the vector of inputs into a single value and $G(x)$ is a potentially vector valued function (i.e., it maps n inputs into m results which are potentially constrained). In order to solve the constrained problem we start by assuming that we have a value that solves the nonlinear constraints and form a linear approximation around that point

$$G(x) \approx G(x^*) + \nabla_x G(x^*) dx$$

$$\nabla_x G(x) = \begin{pmatrix} \frac{\partial G_1(x)}{\partial x_1} & \frac{\partial G_1(x)}{\partial x_2} & \dots & \frac{\partial G_1(x)}{\partial x_n} \\ \frac{\partial G_2(x)}{\partial x_1} & \frac{\partial G_2(x)}{\partial x_2} & \dots & \frac{\partial G_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_m(x)}{\partial x_1} & \frac{\partial G_m(x)}{\partial x_2} & \dots & \frac{\partial G_m(x)}{\partial x_n} \end{pmatrix} \quad (2.21)$$

Given the Taylor series expansion of the vector equation, we can see that starting from a feasible point $\bar{x}$ and stepping to another feasible point $\hat{x}$ involves solving the equation

$$G(\bar{x}) - G(\hat{x}) = \nabla_x G(\hat{x}) dx = 0 \quad (2.22)$$

thus any dx in the null space of $\nabla_x G(\hat{x})$ maintains feasibility.

Simplifying this in the current case assume that $\nabla_x G(\hat{x}) dx$ can be written as

$$\begin{bmatrix} G_1 & G_2 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = 0 \quad (2.23)$$

where G_1 and G_2 are the partitioned values from the Jacobian matrix evaluated at $\hat{x}$. This partitioning would be such that G_1 is an m by m matrix where m is the number of equations in the vector valued function. Thus, G_2 is a m by $n - m$ matrix. This partitioning gives us the matrix equation

$$G_1 dx_1 + G_2 dx_2 = 0 \quad (2.24)$$

Notice that this equation can be solved for dx_1 as

$$dx_1 = -G_1^{-1} G_2 dx_2 \quad (2.25)$$

which gives the change in x_1 associated with a change in x_2 which maintains feasibility in the constraints.

To make this development a little more concrete, consider a slight reformulation of the joint production problem presented in Equation 2.15

$$\begin{aligned} \max_{y_1^*, y_2^*, x_1, x_2, x_3} \quad & 3.75y_1^* + 4.65y_2^* - 1.50x_1 - 2.00x_2 - 1.25x_3 \\ & y_1^* = x_1^{0.15} x_2^{0.25} x_3^{0.55} \\ & y_2^* = x_1^{0.30} x_2^{0.35} x_3^{0.30} \end{aligned} \quad (2.26)$$

The Jacobian or matrix of first derivatives for the two constraints can be written as

$$\nabla_x G(x) = \begin{bmatrix} \frac{0.15x_2^{0.25} x_3^{0.55}}{x_1^{0.85}} & \frac{0.25x_1^{0.15} x_3^{0.55}}{x_2^{0.75}} & \frac{0.55x_1^{0.15} x_2^{0.25}}{x_3^{0.45}} \\ \frac{0.30x_2^{0.35} x_3^{0.30}}{x_1^{0.70}} & \frac{0.35x_1^{0.30} x_3^{0.30}}{x_2^{0.65}} & \frac{0.30x_1^{0.30} x_2^{0.25}}{x_3^{0.70}} \end{bmatrix} \quad (2.27)$$

Substituting the solution from Equation 2.17 yields $y_1^* = 2.53188$ and $y_2^* = 2.31908$ with a Jacobian matrix of

$$\nabla_x G(x^*) = \begin{bmatrix} 0.21596 & 0.356378 & 0.336703 \\ 0.349128 & 0.403293 & 0.148452 \end{bmatrix}. \quad (2.28)$$

To maintain feasibility we divide the Jacobian into two parts

$$G_1 = \begin{bmatrix} 0.21596 & 0.356378 \\ 0.349128 & 0.403293 \end{bmatrix} \quad G_2 = \begin{bmatrix} 0.336703 \\ 0.148452 \end{bmatrix} \quad (2.29)$$

The change in value of x_3 that maintains feasibility can then be computed as

$$dx_3 = - \begin{bmatrix} 0.21596 & 0.356378 \\ 0.349128 & 0.403293 \end{bmatrix}^{-1} \begin{bmatrix} 0.336703 \\ 0.148452 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}. \quad (2.30)$$

Returning to the basic optimization problem, we develop our standard Taylor series expansion of the objective function

$$f(x) - f(x^*) = \nabla_x f(x^*) dx + \frac{1}{2} dx' \nabla_{xx}^2 f(x^*) dx \leq 0 \quad (2.31)$$

Partitioning this expression in the same manner as the constraint matrix, we have

$$\begin{bmatrix} f_1 & f_2 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dx_1 & dx_2 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \leq 0 \quad (2.32)$$

substituting for the feasibility constraint yields

$$\begin{aligned} & \begin{bmatrix} f_1 & f_2 \end{bmatrix} \begin{bmatrix} -G_1^{-1}G_2dx_2 \\ dx_2 \end{bmatrix} + \\ & \frac{1}{2} \begin{bmatrix} -dx_2'G_2'G_1^{-1'} & dx_2 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} -G_1^{-1}G_2dx_2 \\ dx_2 \end{bmatrix} \leq 0 \end{aligned} \quad (2.33)$$

The first necessary condition for this problem can then be defined as

$$\begin{aligned} -f_1G_1^{-1}G_2dx_2 + f_2dx_2 &= 0 \\ [f_2 - f_1G_1^{-1}G_2]dx_2 &= 0 \\ \|f_2 - f_1G_1^{-1}G_2\| &= 0 \end{aligned} \quad (2.34)$$

Intuitively, this condition implies that the gradient projected along a feasible path must be equal to zero at optimum. The second order necessary conditions are then defined by

$$\begin{aligned} & \begin{bmatrix} -dx_2'G_2'G_1^{-1'} & dx_2' \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} G_1^{-1}G_2dx_2 \\ dx_2 \end{bmatrix} = \\ & \begin{bmatrix} dx_2'G_2'G_1^{-1}F_{11} + dx_2'F_{21} & dx_2'G_2'G_1^{-1'}F_{12} + dx_2'F_{22} \end{bmatrix} \begin{bmatrix} G_1^{-1}G_2dx_2 \\ dx_2 \end{bmatrix} \\ & \begin{bmatrix} dx_2'G_2'G_1^{-1'}F_{11}G_1^{-1}G_2dx_2 + dx_2'F_{21}G_1^{-1}G_2dx_2 + \\ dx_2'G_2'G_1^{-1'}F_{12}dx_2 + dx_2'F_{22}dx_2 \end{bmatrix} \\ & dx_2' \left[G_2'G_1^{-1'}F_{11}G_1^{-1}G_2 + F_{21}G_1^{-1}G_2 + G_2'G_1^{-1'}F_{12} + F_{22} \right] dx_2 \end{aligned} \quad (2.35)$$

2.4.3 Imposing Linear Equality Constraints

The general programming problem presented in Chapter 1 can be written

$$\begin{aligned} & \max_x f(x) \\ & \text{s.t } G(x) = b \end{aligned} \quad (2.36)$$

can be restructured into the linearly constrained equality problem

$$\begin{aligned} \max_x f(x) \\ \text{s.t. } Ax = b \end{aligned} \quad (2.37)$$

Mathematically one such problem is the portfolio choice problem where an individual minimizes the variance of an investment portfolio subject to an income constraint

$$\begin{aligned} \min_x f(x) = x' \Omega x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\ \omega_{21} & \omega_{22} & \omega_{23} & \omega_{24} \\ \omega_{31} & \omega_{32} & \omega_{33} & \omega_{34} \\ \omega_{41} & \omega_{42} & \omega_{43} & \omega_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \\ = \omega_{11}x_1^2 + 2\omega_{12}x_1x_2 + 2\omega_{13}x_1x_3 + 2\omega_{14}x_1x_4 + \omega_{22}x_2^2 + \\ 2\omega_{23}x_2x_3 + 2\omega_{24}x_2x_4 + \omega_{33}x_3^2 + 2\omega_{34}x_3x_4 + \omega_{44}x_4^2 \\ \text{s.t. } \mu_1x_1 + \mu_2x_2 + \mu_3x_3 + \mu_4x_4 = Y^* \end{aligned} \quad (2.38)$$

Mathematically, solving this system is based on the null-space of the income constraint. In this case, setting the level of x_2 through x_4 determines the level of x_1

$$x_1 = \frac{Y^*}{\mu_1} - \frac{\mu_2}{\mu_1}x_2 - \frac{\mu_3}{\mu_1}x_3 - \frac{\mu_4}{\mu_1}x_4. \quad (2.39)$$

We could substitute for this equality into the original objective function

$$\begin{aligned} f^*(x) = \omega_{11} \left(\frac{Y^*}{\mu_1} - \frac{\mu_2}{\mu_1}x_2 - \frac{\mu_3}{\mu_1}x_3 - \frac{\mu_4}{\mu_1}x_4 \right)^2 + \\ 2\omega_{12}x_2 \left(\frac{Y^*}{\mu_1} - \frac{\mu_2}{\mu_1}x_2 - \frac{\mu_3}{\mu_1}x_3 - \frac{\mu_4}{\mu_1}x_4 \right) + \\ 2\omega_{13}x_3 \left(\frac{Y^*}{\mu_1} - \frac{\mu_2}{\mu_1}x_2 - \frac{\mu_3}{\mu_1}x_3 - \frac{\mu_4}{\mu_1}x_4 \right) + \\ 2\omega_{14}x_4 \left(\frac{Y^*}{\mu_1} - \frac{\mu_2}{\mu_1}x_2 - \frac{\mu_3}{\mu_1}x_3 - \frac{\mu_4}{\mu_1}x_4 \right) + \\ \omega_{22}x_2^2 + 2\omega_{23}x_2x_3 + 2\omega_{24}x_2x_4 + \omega_{33}x_3^2 + 2\omega_{34}x_3x_4 + \omega_{44}x_4^2. \end{aligned} \quad (2.40)$$

reducing the multivariate constrained problem to a multivariate unconstrained problem. However, this scenario becomes impractical as the number of equations and variables grow.

Another way to conceive the problem is by the formal definition of the null space. What we want to do is to develop a formalized way of making

the statement that $p = (\bar{x} - \hat{x})$ implies that $A\bar{x} = b$ and $A\hat{x} = b$. Thus, $A\bar{x} = A\hat{x}$ implies that $A(\bar{x} - \hat{x}) = 0$ or that $Ap = 0$. The collection of all such vectors is referred to as the null space. Specifically, let the columns of matrix Z form a basis; then $AZ = 0$ defines the null space.

In this case, we need to modify our concept of the neighborhood. Previously, we were interested in any projection around x^* , thus $x = x^* + p$. Now, we are only interested in those points that maintain feasibility. To accomplish this, we begin with a projection (or perturbation) in the same dimension as the null space. This perturbation is then mapped into the general space using the null space basis. Thus, the set of feasible x s becomes $x = x^* + Zp$. The second order expansion of the function then becomes

$$f(x) = f(x^*) + \nabla_x f(x^*) Zp + \frac{1}{2} p' Z' \nabla_{xx}^2 f(x^*) Zp. \quad (2.41)$$

Again using the mechanics for a maximum, we have that

$$f(x) - f(x^*) = \nabla_x f(x^*) Zp + \frac{1}{2} p' Z' \nabla_{xx}^2 f(x^*) Zp \leq 0 \quad (2.42)$$

for any p . Hence, we derive the first necessary condition for a maximum that

$$\nabla_x f(x^*) Z = 0 \quad (2.43)$$

which states that the projected gradient is equal to zero at optimal. Similarly, the second order necessary condition for an optimal becomes

$$Z' \nabla_{xx}^2 f(x^*) Z \text{ is negative semi-definite} \quad (2.44)$$

Summary of Optimality Conditions for Unconstrained and Linearly Constrained

- Necessary Conditions
 - Unconstrained Optimization
 - * $\|\nabla_x f(x^*)\| = 0$ or x^* is a stationary point.
 - * $\nabla_{xx}^2 f(x^*)$ is negative semi-definite.
 - Linearly Constrained Optimization

- * $Ax^* = b$ or the point is feasible.
 - * $\|Z'\nabla_x f(x^*)\| = 0$ or the projected gradient is equal to zero. This point can be equivalently defined by $\nabla_x f(x^*) = A\lambda^*$ where λ^* is the vector of shadow values.
 - * $Z'\nabla_{xx}^2 Z$ is negative semi-definite
- Sufficient Conditions
 - Unconstrained Optimization
 - * $\|\nabla_x f(x^*)\| = 0$ or x^* is a stationary point.
 - * $\nabla_{xx}^2 f(x^*)$ is negative definite.
 - Linearly Constrained Optimization
 - * $Ax^* = b$ or the point is feasible.
 - * $\|Z'\nabla_x f(x^*)\| = 0$ or the projected gradient is equal to zero. This point can be equivalently defined by $\nabla_x f(x^*) = A\lambda^*$ where λ^* is the vector of shadow values.
 - * $Z'\nabla_{xx}^2 Z$ is negative definite

2.5 An Applied Example

Using data from Levy and Sarnat, the variance matrix in the above problem can be set as

$$\Omega = \begin{pmatrix} 924.41 & 458.52 & 202.22 & 135.22 \\ 458.52 & 761.29 & 452.99 & 72.55 \\ 202.22 & 452.99 & 490.11 & 109.09 \\ 135.22 & 72.25 & 109.09 & 284.17 \end{pmatrix} \quad (2.45)$$

Next, we impose two constraints. The first constraint is the income constraint and the second is the portfolio balance constraint

$$\begin{aligned} 8.119x_1 + 11.366x_2 + 6.298x_3 + 8.014x_4 &= 7.00 \\ x_1 + x_2 + x_3 + x_4 &= 1.00 \end{aligned} \quad (2.46)$$

Using MINOS 5.6 (as described in Appendix B) the optimal portfolio of assets can be computed as

$$x = \begin{pmatrix} 0.11613 \\ -0.10666 \\ 0.39508 \\ 0.59545 \end{pmatrix} \quad (2.47)$$

Summing the vector of xs we see that the sum is 1.0 and the expected return is 7.0. The necessary conditions for optimality then become the projected gradient and hessian matrices. The null space matrix for the constraint set is

$$Z = \begin{pmatrix} -0.86287 & 0.00000 \\ 0.22195 & -0.271782 \\ 0.340534 & -0.530893 \\ 0.300383 & 0.802675 \end{pmatrix} \quad (2.48)$$

The gradient vector evaluated at the solution is

$$\nabla_x f(x) = \begin{pmatrix} 437.714 \\ 388.076 \\ 467.519 \\ 440.616 \end{pmatrix} \quad (2.49)$$

The projected gradient then becomes

$$Z' \nabla_x f(x) = (0.0036136 \ -0.00288491) \quad (2.50)$$

2.6 Linear Inequality

As previously developed, the first-order necessary conditions for a linearly constrained equality can be written as

$$Z' \nabla_x f(x) = 0 \text{ or } \nabla_x f(x) = A' \lambda \quad (2.51)$$

where λ is a vector of Lagrange Multipliers. Another way to look at the problem is that the Lagrange multipliers are that set of constants which equate the gradient with the linear constraints.

Proof. Starting with a two constraint model

$$L = f(x) + \lambda_1 (b_1 - a_1 x) + \lambda_2 (b_2 - a_2 x) \quad (2.52)$$

where a_1 is a row vector (conformable to x) and a_2 is a column vector. The expression in Equation 2.52 can be rewritten as

$$L = f(x) - \lambda Ax \quad (2.53)$$

in this example A is a $2 \times n$ matrix (where there are n variables in x). Thus, Ax is a 2×1 vector and λ is a 1×2 vector.

Forming the gradient vector for the Lagrangian (i.e., taking the derivatives of Equation 2.52 yields

$$\nabla_x L = \nabla_x f(x) - \lambda_1 a_1 - \lambda_2 a_2 = 0 \text{ or } \nabla_x f(x) = A' \lambda. \quad (2.54)$$

□

The general concept of the lagrange multiplier and the null space matrix then work together.

Another concept that I want to discuss is the fact that a equality constrained problem can have lagrange multipliers of either side. Intuitively, the lagrange multiplier is the change in the objective function associated with a one unit increase in the right hand side of the constraint. Therefore, in a maximization problem: (1) If the constraint cuts the frontier below the global maximum (assuming that the objective function resembles a quadratic function), the lagrange multiplier will take on a positive value implying that an increase in the right hand side of the constraint will increase the objective function value. And (2) if the constraint cuts the frontier above the global maximum (again assuming that the objective function resembles a quadratic function), the lagrange multiplier will take on a negative value implying that an increase in the right hand side of the constraint will cause the objective function to decline.

A numerical example based on the Levy and Sarnat portfolio problem. The formulation above minimizes the variance of a portfolio with a return of 7 percent. The optimal portfolio was

$$x = \begin{pmatrix} 0.11613 \\ -0.10666 \\ 0.39508 \\ 0.59546 \end{pmatrix} \quad (2.55)$$

The constraint matrix for this problem was

$$A = \begin{pmatrix} 8.199 & 11.366 & 6.298 & 8.014 \\ 1.0 & 1.0 & 1.0 & 1.0 \end{pmatrix} \quad (2.56)$$

Forming the lagrange expression, we will try to find λ_1 and λ_2 such that

$$\nabla_x f(x^*) = \begin{pmatrix} 437.712 \\ 388.069 \\ 467.511 \\ 440.613 \end{pmatrix} = \begin{pmatrix} 8.199 & 1.0 \\ 11.366 & 1.0 \\ 6.298 & 1.0 \\ 8.014 & 1.0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \quad (2.57)$$

The two shadow values can be found by solving the system

$$\begin{pmatrix} 8.199 & 1.0 & 436.712 \\ 11.366 & 1.0 & 388.069 \\ 6.298 & 1.0 & 467.511 \\ 8.014 & 1.0 & 440.613 \end{pmatrix} \quad (2.58)$$

which yields a row reduced form of

$$\begin{pmatrix} 1 & 0 & -15.675 \\ 0 & 1 & 566.212 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (2.59)$$

One cautionary note is that using automatic row-reduction procedures in Gauss or Mathematica may reduce the problem beyond the first two row operations due to rounding problems. I would recommend manual reductions.

The conditions for inequality constraints are built from the concepts in the linear equality constraints. Specifically, the general form of the linearly constrained inequality programming (LIP) problem is stated

$$\begin{aligned} \max_x & f(x) \\ \text{s.t. } & Ax \leq b \end{aligned} \quad (2.60)$$

Again, we will use the null space, but in applying the null space we will divide the set of linear constraints into active or binding and nonactive or nonbinding constraints (depicted graphically in Figure 2.1). The binding constraints must satisfy all the properties that an equality must satisfy, but a nonbinding constraint is omitted from the construction of the null space.

Mathematically, the necessary conditions for a linear inequality constrained problem are

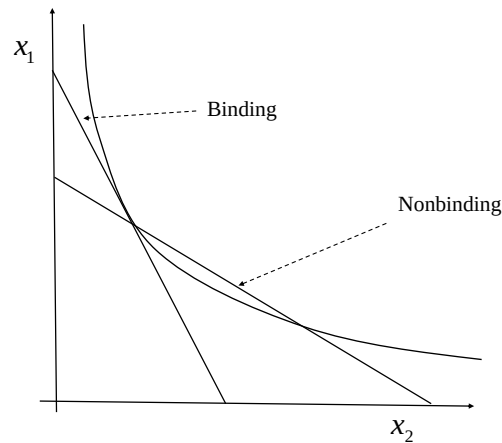


Fig. 2.1 Binding and Nonbinding Constraints

- $Ax \leq b$
- $Z'\nabla_x f(x) = 0$ or $\nabla_x f(x) = A'\lambda$
- $\lambda \geq 0$
- $Z'\nabla_{xx}^2 f(x)Z$ is negative semi-definite

where Z (the null space matrix is defined for only the active or binding constraints of A).

Given these conditions we see that the portfolio problem discussed above is not consistent with income greater than or equal to 7 percent rather than exactly equal 7 percent. Specifically, $\lambda_1 = -15.675$. Thus, the income constraint is constraining income below the global minimum for variance.

Another example developed in the mathematica notebook Constrained2.ma is the utility maximization problem where the consumer maximizes utility defined as:

$$\begin{aligned} & \max_x (4 \ 3 \ 1 \ 9) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} - \\ & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} \begin{pmatrix} 1 & 0.025 & -0.0125 & 0.0166 \\ 0.025 & 0.750 & 0.0005 & -0.001 \\ -0.0125 & 0.0005 & 0.25 & 0.045 \\ 0.0166 & -0.001 & 0.045 & 1.25 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \end{aligned} \quad (2.61)$$

Subject to the income constraint

$$(3 \ 2 \ 4 \ 9) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \leq Y \quad (2.62)$$

We see that the unconstrained problem has a global maximum in the positive quadrant. The exact maximum is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1.91157 \\ 1.94001 \\ 1.45741 \\ 3.52385 \end{pmatrix} \quad (2.63)$$

The required level of income to meet this global maximum is 33.0636.

Thus, we want to evaluate two scenarios. Under the first scenario the level of income is set at 40.0 which is beyond the level required for the global maximum. Under this scenario we solve

$$\begin{aligned} & \max_x f(x) \\ & \text{s.t. } (3 \ 2 \ 4 \ 5) x = 40 \end{aligned} \quad (2.64)$$

which yields an optimal solution of $\lambda = -0.14383$. Following the classical necessary conditions for optimization, we see that this point fails even though the projected gradient is zero because the sign on the Lagrange multiplier is wrong. The utility maximizing point in this case is actually the global maximum.

The second scenario then involves constraining y to be less than 30. In this case, the optimum solution becomes

$$x = \begin{pmatrix} 1.81084 \\ 1.85618 \\ 0.94317 \\ 3.41237 \end{pmatrix} \quad (2.65)$$

with $\lambda = 0.065537$.

Under this scenario, it is possible to show that (1) the constraint set is met, (2) the projected gradients vanish, (3) the Lagrange multipliers are greater than or equal to zero, and (4) the projected Hessian is negative definite. Thus, this solution meets the first and second order necessary and sufficient conditions for a maximum.

2.7 Necessary and Sufficient Conditions for Nonlinear Constraints

2.7.1 Simple Constraint

To begin our discussion, let us assume that we want to maximize a quadratic utility function subject to an elliptical constraint. Abstracting away from the objective function I want to develop the meaning of the first order Taylor expansion of this elliptical constraint at a given point. Let the elliptical constraint be

$$g(x) = 10 - 3x_1^2 - 2x_2^2 \quad (2.66)$$

This surface is plotted in Figure 2.2. To generate this surface we solve for the implicit function of x_1 in terms of x_2

$$g(x) \Rightarrow x_1 = \frac{\sqrt{6}}{3} \sqrt{5 - x_2^2} \quad (2.67)$$

Next, we want to linearize this function at a particular point, say $x_2 = .75$. Solving the implicit function at this point yields $x_1 = 1.71998$ as graphed in Figure 2.3.

The Taylor series expansion of this constraint is then defined as

$$G(x) \approx G(x^*) + \nabla_x G(x^*) dx \quad (2.68)$$

The Jacobian of this expansion is defined as

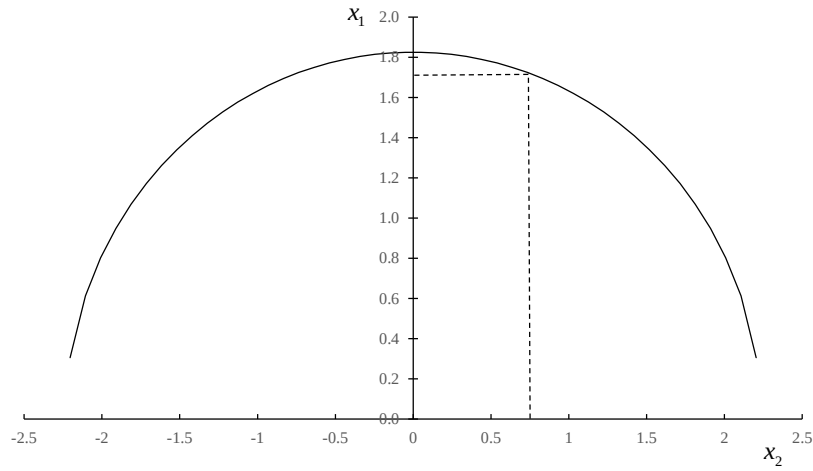


Fig. 2.2 Nonlinear Constraint

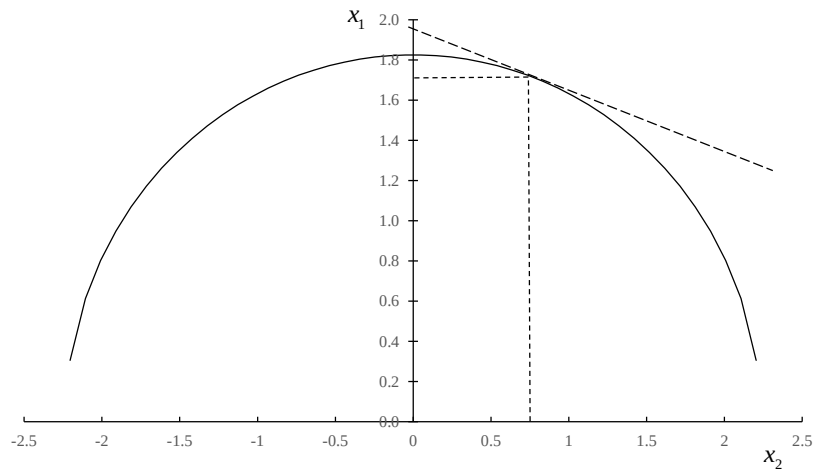


Fig. 2.3 Linear Approximation

$$\nabla_x G(x) = \begin{pmatrix} \frac{\partial G_1(x)}{\partial x_1} & \frac{\partial G_1(x)}{\partial x_2} & \dots & \frac{\partial G_1(x)}{\partial x_n} \\ \frac{\partial G_2(x)}{\partial x_1} & \frac{\partial G_2(x)}{\partial x_2} & \dots & \frac{\partial G_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_m(x)}{\partial x_1} & \frac{\partial G_m(x)}{\partial x_2} & \dots & \frac{\partial G_m(x)}{\partial x_n} \end{pmatrix} \quad (2.69)$$

is the cornerstone of the linear approximation. For the current problem, the Jacobian at the point of approximation is

$$\nabla_x G(x) = (-10.3199 \ -3.00) \quad (2.70)$$

The linear relationship around that point then becomes

$$G(x) - G(x^*) = (-10.3199 \ -3.0000) \begin{pmatrix} x_1 - 1.71998 \\ 0.75 \end{pmatrix} \quad (2.71)$$

$$G(x) - G(x^*) = 20 - 10.3199x_1 - 3.000x_2$$

Optimality Conditions

- $G(x) = 0$.
- $\|Z(x)' \nabla_x f(x)\| = 0$ or $\nabla_x f(x) = A(x)' \lambda$.
- $Z(x)' W(x, \lambda) Z(x)$ is positive semi-definite for the necessary conditions (and positive definite for the sufficient conditions).

where

$$W(x, \lambda) = \nabla_{xx}^2 f(x) - \sum_{i=1}^m \lambda_i \nabla_{xx}^2 G_i(x) \quad (2.72)$$

Chapter 3

Algorithms for Optimization

3.1 Univariate Newtons's Method

As a starting point consider the problem of finding the maximum of a quadratic utility function

$$U(x) = 5x^2 - 4x + 2 \quad (3.1)$$

which has a minimum determined by setting its first derivative equal to zero

$$\begin{aligned} \frac{\partial U(x)}{\partial x} &= 10x - 4 = 0 \\ x &= \frac{2}{5} \end{aligned} \quad (3.2)$$

In addition, straightforward transformations such as

$$U(x) = e^{5x^2 - 4x + 2} \quad (3.3)$$

offer little additional complexity

$$\begin{aligned} \frac{\partial U(x)}{\partial x} &= U(x) [10x - 4] = 0 \\ x &= \frac{2}{5} \end{aligned} \quad (3.4)$$

Thus, the functions presented in Equations 3.1 and 3.3 are not complex enough to be interesting functions to discuss algorithms. Another possibility is a rational function

$$f(x) = \frac{5x^2 - 4x + 2}{x + 10} \quad (3.5)$$

which is depicted in Figure 3.1. However, these functions typically have bizarre discontinuities as depicted in a plot of a broader range for the same function in Equation 3.1. This plot shows the vertical asymptote at $x = -10$.

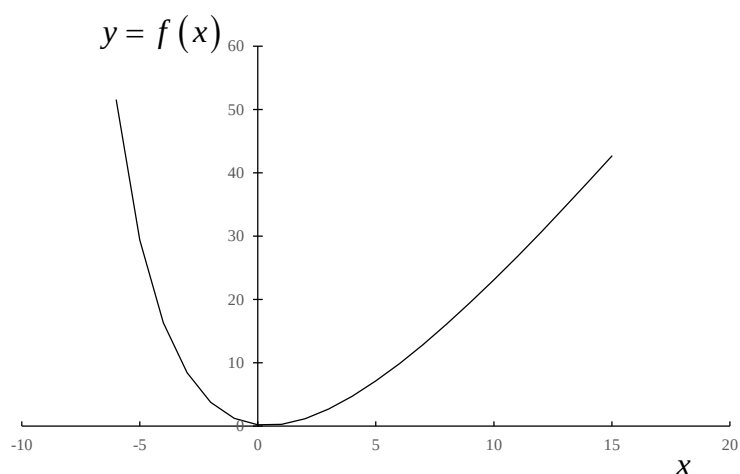


Fig. 3.1 Graph of Rational Function

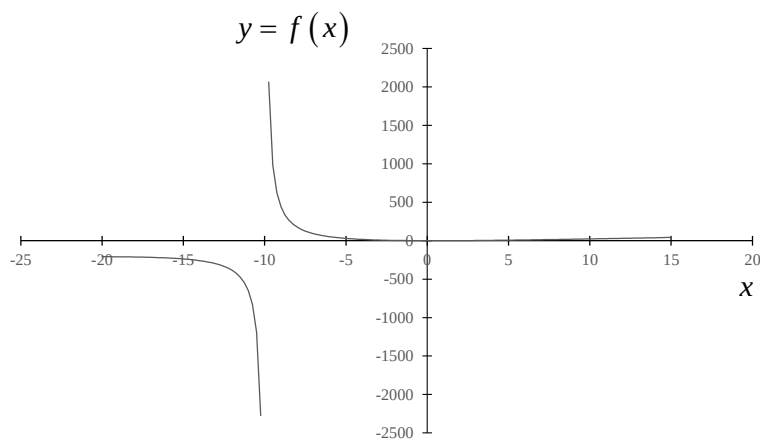


Fig. 3.2 Extended Graph of Rational Function

If we restrict our attention to the range of x 's such that x is greater than -10 the problem becomes more tractable. Taking the first-order condition

$$\begin{aligned}\frac{\partial f(x)}{\partial x} &= \frac{10x - 4}{10 - x} - \frac{5x^2 - 4x + 2}{(x + 10)^2} = 0 \\ x &= \frac{-100 + 2\sqrt{2710}}{10} \\ x &= 0.411532\end{aligned}\tag{3.6}$$

Most of the numeric optimization procedures involve numerically finding the zero of the derivative. The gradient of the rational over the range $x \in (-10, \infty]$ is presented in Figure 3.3.

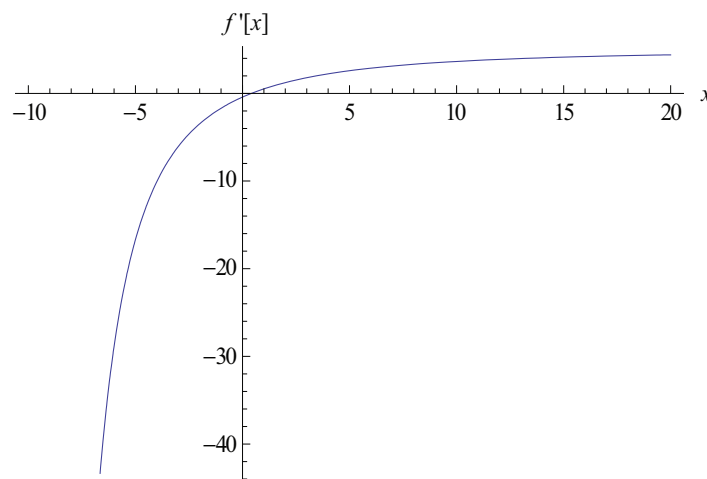


Fig. 3.3 Derivative of Rational Function

3.1.1 Method of Bisection

The method of bisection involves shrinking the interval. Specifically, suppose that the original interval was $[-8, 20]$. Evaluating the function at each point $f[-8] = 16.111$ and $f[20] = 64.067$. The midpoint of the interval is $x = 6$. At $x = 6$, the value of the derivative is $f[6] = 2.88281$. Therefore, we throw away to higher bisection yielding an interval of $[-8, 6]$ (we throw away the bisection with the same sign on the derivative). Next, taking the midpoint of the new interval is -1 with a derivative value of -1.69136. At

this iteration, we throw out the lower bisection yielding an interval of $[-1, 6]$. Table 3.1 presents a complete sequence of bisections.

3.1.2 Newton's Method

The theoretical foundation of Newton's method involves inscribing triangles inside the function. Mathematically

$$\begin{aligned}\Delta y &\approx \frac{\partial f(x)}{\partial x} \Delta x \\ \Delta x &= \frac{\Delta y}{\frac{\partial f(x)}{\partial x}} \\ x_{t+1} &= x_t - \frac{f(x)}{\frac{\partial f(x)}{\partial x}}\end{aligned}\tag{3.7}$$

This Newton step is depicted graphically in Figure 3.4. Table 3.2 presents the steps for the Newton search algorithm for the derivative of the rational function presented in Equation 3.6.

3.2 Finding the Multivariate Maximum

The basic difference between univariate and multivariate optimization is the number of equations which we want to solve simultaneously for zero. In the multivariate case we want to solve

$$\nabla_x f(x) = 0\tag{3.8}$$

where x is an n element vector, so we want to solve for n equations equal to zero. Appealing again to the second order Taylor series expansion

$$\begin{aligned}\nabla_x f(x) &= \nabla_x f(x^*) + \nabla_{xx}^2 f(x^*) (x - x^*) = 0 \\ \Rightarrow (x - x^*) &= (\nabla_{xx}^2 f(x^*))^{-1} \nabla_x f(x)\end{aligned}\tag{3.9}$$

In this case, we take a linear approximation of the gradient function (i.e., the $n \times 1$ vector of derivatives $\nabla_x f(x)$). This approximation following the first-order Taylor approximation is a function of the original gradient function at the point of approximation ($\nabla_x f(x^*)$) and the derivative (matrix) of the gradient (i.e., the hessian matrix of the objective function $\nabla_{xx}^2 f(x^*)$).

Iteration	Lower Value	Lower Derivative	Upper Value	Upper Derivative	Bisection Value	Bisection Derivative	Replace
0	-8.0000	-130.5000	20.0000	4.3978	6.0000	2.8828	Upper
1	-8.0000	-130.5000	6.0000	2.8828	-1.0000	-1.6914	Lower
2	-1.0000	-1.6914	6.0000	2.8828	2.5000	1.5312	Upper
3	-1.0000	-1.6914	2.5000	1.5312	0.7500	0.3099	Upper
4	-1.0000	-1.6914	0.7500	0.3099	-0.1250	-0.5581	Lower
5	-0.1250	-0.5581	0.7500	0.3099	0.3125	-0.0965	Lower
6	0.3125	-0.0965	0.7500	0.3099	0.5313	0.1130	Upper
7	0.3125	-0.0965	0.5313	0.1130	0.4219	0.0099	Upper
8	0.3125	-0.0965	0.4219	0.0099	0.3672	-0.0429	Lower
9	0.3672	-0.0429	0.4219	0.0099	0.3945	-0.0164	Lower
10	0.3945	-0.0164	0.4219	0.0099	0.4082	-0.0032	Lower
11	0.4082	-0.0032	0.4219	0.0099	0.4150	0.0034	Upper
12	0.4082	-0.0032	0.4150	0.0034	0.4116	0.0001	Upper
13	0.4082	-0.0032	0.4116	0.0001	0.4099	-0.0016	Lower
14	0.4099	-0.0016	0.4116	0.0001	0.4108	-0.0007	Lower
15	0.4108	-0.0007	0.4116	0.0001	0.4112	-0.0003	Lower

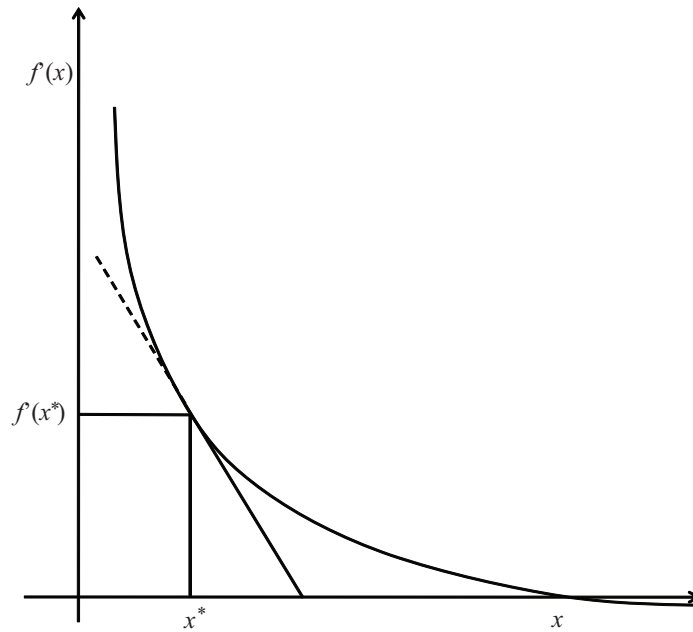


Fig. 3.4 Newton Step

Search Point	Function Value	Derivative	Step
-8.0000	-130.5000	135.5000	0.9631
-7.0369	-56.7315	41.6668	1.3616
-5.6754	-23.9799	13.4022	1.7893
-3.8861	-9.4998	4.7432	2.0028
-1.8833	-3.2269	2.0272	1.5919
-0.2914	-0.7503	1.1846	0.6334
0.3419	-0.0675	0.9800	0.0689
0.4108	-0.0007	0.9607	0.0007
0.4115	0.0000	0.9605	0.0000

Substituting x_{t+1} for x and x_t for x^* we are left with an estimate of the value of x for the next iteration such that the value of the derivative is equal to zero

$$x_{t+1} = x_t - (\nabla_{xx}^2 f(x_t))^{-1} \nabla_x f(x_t) \quad (3.10)$$

3.2.1 A Simple Example

As a first problem, consider a Cobb-Douglas utility function with a budget constraint imposed

$$\begin{aligned} \max_x & x_1^{0.2} x_2^{0.3} x_3^{0.4} x_4^{0.1} \\ \text{s.t. } & x_1 + 2x_2 + 3x_3 + x_4 = 100 \end{aligned} \quad (3.11)$$

By substituting the constraint for Equation 3.11 into its objective function the unconstrained maximization problem becomes

$$\max_x (100 - 2x_2 - 3x_3 - x_4)^{0.2} x_2^{0.3} x_3^{0.4} x_4^{0.1} \quad (3.12)$$

Starting with the point $x(1 \ 1 \ 1)$ (note that we now only have three variables because the value of x_1 is determined by the constraint $x_1 = 100 - 2x_2 - 3x_3 - x_4$)

$$\nabla_x f(x) = (0.7337 \ 0.9766 \ 0.2428)$$

$$\nabla_{xx}^2 f(x) \begin{pmatrix} -0.5275 & 0.2885 & 0.0717 \\ 0.2885 & -0.6085 & 0.0954 \\ 0.0717 & 0.0954 & -0.2244 \end{pmatrix}$$

$$x_{t+1} = (1 \ 1 \ 1) - \nabla_x f(x_t) (\nabla_{xx}^2 f(x_t))^{-1} = (5.3559 \ 5.3479 \ 5.3226) \quad (3.13)$$

Need to follow through with the numbers (i.e., the point that the new solution is better than the old solution and that iteratively applying the procedure will converge to an optimum)

3.2.2 Prototype Algorithm

The idea that the rules defined by the Newton-Raphson approach above applied multiple times yields a value the x_t vector which is arbitrarily close to the optimal point defines an Algorithm. More specifically, an algorithm is a set of rules that can be mechanically applied to yield an approximation of the optimal value of the function (either maximum or minimum). Following the example of [Gill *et al.* (1981)] we will define a prototype algorithm for solving the unconstrained optimization algorithm [U] based on the Newton Raphson process. In this definition we assume that we start with a current estimate of $x^*(x_t)$ and attempt to find a point that minimizes the objective function $f(x)$

- U1 *Test for convergence* If the current point (x_t) satisfies the condition for convergence the algorithm terminates with x_t . In the case of the unconstrained problem or convergence condition is $|\nabla_x f(x_t)| = 0$. If x_t satisfies this condition $f(x_t)$ is a local minimum.
- U2 *Compute a search direction* If the current point is not optimum then what direct do we want to move to attempt to find an optimum. In an univariate application the search direction is either to the right (a higher value of x) or to the left (a lower value of x) on the number line. In the multivariate case we can proceed in an infinite number of directions (i.e., in the two variable case an infinite number of x_1/x_2). The direction of the search is a $n \times 1$ projection vector p_t which defines the direction of the search. In the Newton-Raphson approach $\nabla_x f(x_t) (\nabla_{xx}^2 f(x_t))^{-1}$ defines the direction of the search.
- U3 *Compute a step length* Given that we know the direction of the search the next step is to compute how far to step a_t . In Equation 3.13 our implicit step length was 1. However, if the function is concave so that the derivative is decreasing at a decreasing rate we know that the zero implies a larger step than the linear approximation. The question is how much farther could we go? An appropriately selected step length reduces the number of iterations required for minimization. However, a step length that is too large could potentially increase the number of iterations required. At this stage, it is sufficient to state that the step length is chosen to reduce the value of the objective function $f(x_t + a_t p_t) < f(x_t)$.
- U4 *Update the estimate of the minimum* Set $x_{t+1} = x_t + a_t p_t$ and go back to step U1

Given the steps to the prototype algorithm, I want to develop a sample problem that we can compare the various algorithms against. Notebook Algorithm3.ma includes the numeric problem from Equation 3.11. The desktop CPU solves this sequence (not optimally specified in Mathematica) in 8 iterations taking 0.374 CPU seconds.

3.3 Quasi-Newton Methods for Optimization

The Newton-Raphson methodology can be used in U2 in the prototype algorithm. Specifically, the search direction can be determined by

$$p_t = -(\nabla_{xx}^2 f(x_t))^{-1} \nabla_x f(x_t) \quad (3.14)$$

Quasi-Newton algorithms involve an approximation to the Hessian matrix. Replacing the Hessian may be advantageous because the true Hessian is analytically or numerically complex.

3.3.1 Steepest Descent

For example, we could replace the Hessian matrix in Equation 3.14 with the negative of the identity matrix (i.e., a constant negative definite matrix) and the applying the process will yield a new x_{t+1} that is closer to the optimum than our beginning estimate. In this case the search direction would be

$$p_t = -(-I(n)) \nabla_x f(x_t) \quad (3.15)$$

where $I(n)$ is the identity matrix conformable with the gradient vector. This replacement is referred to as the steepest descent method. In our sample problem, this methodology requires 990 iterations and 0.141 seconds. This result highlights some interesting features regarding the Quasi-Newton methods. The steepest descent method requires more overall iterations. In this example, the steepest descent method requires 99 times as many iterations as the Newton-Raphson method. However, the time spent on each iteration is reduced. Again, in the current comparison each the steepest descent method requires .123 seconds per iteration while Newton-Raphson requires .030 seconds per iteration. **This section needs updating with the new computers.**

Obviously substituting the identity matrix uses no real information from the Hessian matrix. An alternative to this drastic reduction would be to systematically derive a matrix H_k which uses curvature information akin to the Hessian matrix. The projection could then be derived as

$$p_t = -H_k^{-1} \nabla_x f(x_t) \quad (3.16)$$

3.3.2 Conjugate Gradient Methods

One class of Quasi-Newton methods are the conjugate gradient methods which "build" up information on the Hessian matrix. From our standard starting point, we take a Taylor series expansion around the point $x_t + s_t$

$$\nabla_x f(x_t + s_t) = \nabla_x f(x_t) + \nabla_{xx}^2 f(x_t) s_t \quad (3.17)$$

for some $s_t \in N(x_t, \delta)$. Solving this expression for the term involving the Hessian yields

$$\begin{aligned} \nabla_x f(x_t + s_t) - \nabla_x f(x_t) &= \nabla_{xx}^2 f(x_t) s_t \\ s_t' (\nabla_{xx}^2 f(x_t + s_t) - \nabla_{xx}^2 f(x_t)) &= s_t' \nabla_{xx}^2 f(x_t) s_t. \end{aligned} \quad (3.18)$$

Thus, the change in the gradient contains the same curvature information as the original Hessian.

Defining the change in the gradient as y_t for simplification

$$y_t = \nabla_x f(x_t + s_t) - \nabla_x f(x_t) \Rightarrow y_t = \nabla_{xx}^2 f(x_t) s_t. \quad (3.19)$$

So we want to approximate the information in the Hessian using the some matrix B_{t+1}

$$y_t = B_{t+1} s_t. \quad (3.20)$$

One way to generate B_{t+1} is to start with the current B_t and add new information on the current solution (or to conjugate information based on our previous information and new information in the change of the gradient)

$$\begin{aligned} B_{t+1} &= B_t + uv' \\ y_t &= (B_t + uv') s_t \end{aligned} \quad (3.21)$$

where u and v are $n \times 1$ vectors. Carrying through the vector multiplication

$$u(v' s_t) = y_t - B_t s_t \quad (3.22)$$

where $v' s_t$ is a scalar. Solving for u given v and information from the last iteration

$$u = \frac{1}{v' s_t} (y_t - B_t s_t) \quad (3.23)$$

Substituting the result for u in Equation 3.23 into the expression for B_{t+1} in Equation 3.21 the update is based only on the value of v . In other words, by designing v we could update the conjugate value of the Hessian.

Given this derivation, a logical choice of v would be $y_t + B_t s_t$. This approximation yields an updated Hessian of

$$B_{t+1} = B_t + \frac{1}{(y_t - B_t s_t)' s_t} (y_t - B_t s_t) (y_t - B_t s_t)' \quad (3.24)$$

This needs a little more development. It is a rank-one update what is wrong with that? Also an example walk through may be a good idea This update results in a symmetric approximation to the Hessian. In general, any approximation to the Hessian can be made symmetric by

$$B^{(2)} = \frac{1}{2} (B^{(1)} + B^{(1)'}) \quad (3.25)$$

Substituting such an expression into the general updating procedure results in

$$B_{t+1} = B_t + \frac{1}{v' s_t} ((y_t - B_t s_t) v' + v (y_t - B_t s_t)') - \frac{(y_t - B_t s_t)' s_t}{(v' s_t)^2} v v'. \quad (3.26)$$

The advantage of this formulation is that the approximation of the Hessian matrix will be symmetric for any vector v .

Two prominent conjugate gradient methods are the Davidon-Fletcher-Powell (DFP) update and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update. In the DFP update v is set equal to y_t yielding

$$B_{t+1} = B_t - \frac{1}{s_t' B_t s_t} B_t s_t s_t' B_t + \frac{1}{y_t' s_t} y_t y_t' + (s_t' B_t s_t) w_t w_t' \quad (3.27)$$

$$w_t = \frac{1}{y_t' s_t} y_t - \frac{1}{s_t' B_t s_t} B_t s_t.$$

The BFGS update is then

$$B_{t+1} = B_t - \frac{1}{s_t' B_t s_t} B_t s_t s_t' B_t + \frac{1}{y_t' s_t} y_t y_t'. \quad (3.28)$$

3.3.3 A Numerical Example

Using the previously specified problem and starting with an identity matrix as the original Hessian matrix, each algorithm was used to maximize the utility function. Each of the procedures will have the same first iteration

because they have the same original Hessian matrix and analytical gradient. The initial gradient for all the Quasi-Newton procedures was

$$B_t = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (3.29)$$

compared with an analytical gradient of

$$B_t^* = \begin{pmatrix} -0.5275 & 0.2885 & 0.0718 \\ 0.2885 & -0.6085 & 0.0954 \\ 0.0718 & 0.0954 & -0.2244 \end{pmatrix}. \quad (3.30)$$

This substitution implies a step of

$$s_t = (0.7337 \ 0.9766 \ 0.2428) \quad (3.31)$$

compared with an optimal step of

$$s_t^* (4.3559 \ 4.3476 \ 4.3226) \quad (3.32)$$

In discussing the difference in step, I will focus on two attributes. The first attribute is the relative length of the step (the 2-norm). If the conjugate gradient routines can be made identical with Newton-Raphson with the appropriate step length algorithm, we could then focus on designing step lengths. The second attribute is the direction of the step. As we discussed above one of gains to Newton-Raphson is the refinement of the search direction through the information in the Hessian. The value of this refinement can be seen in the relative gain of the Newton-Raphson over the steepest descent method. Dividing each vector by its 2-norm yields the results depicted in Table 3.3.

3.3.3.1 Rank One Approximation

- Iteration 1

$$B_t = \begin{pmatrix} -0.6690 & 0.3842 & 0.1536 \\ 0.3842 & -0.5540 & 0.1783 \\ 0.1536 & 0.1783 & -0.2280 \end{pmatrix} \quad (3.33)$$

$$B_t^* = \begin{pmatrix} -0.2780 & 0.1276 & 0.0511 \\ 0.1276 & -0.2501 & 0.059 \\ 0.0511 & 0.0592 & -0.2280 \end{pmatrix} \quad (3.34)$$

	Step				
	x_1	x_2	x_3	Length	
<i>Steepest Descent</i>					
Newton-Raphson	4.3559	4.3476	4.3226	7.5207	0.5781 0.5748
Conjugate Gradient	0.7337	0.9766	0.2428	1.2454	0.5891 0.7842 0.1950
<i>Rank One Approximation - Iteration 1</i>					
Newton-Raphson	6.7466	7.6270	4.8254	11.2682	0.5987 0.6769 0.4282
Conjugate Gradient	4.3187	5.0130	2.0042	6.9136	0.6247 0.7251 0.2899
<i>Rank One Approximation - Iteration 2</i>					
Newton-Raphson	10.8541	11.6909	6.0789	17.0717	0.6358 0.6848 0.3561
Conjugate Gradient	7.7298	8.9067	3.7912	12.3876	0.6240 0.7190 0.3060
<i>Powell Symmetric Broyden - Iteration 1</i>					
Newton-Raphson	10.8541	11.6909	6.0789	17.0717	0.6358 0.6848 0.3561
Conjugate Gradient	7.7298	8.9067	3.7912	12.3876	0.6240 0.7190 0.3060
<i>Powell Symmetric Broyden - Iteration 2</i>					
Newton-Raphson	10.8525	11.7000	6.0686	17.0732	0.6356 0.6853 0.3554
Conjugate Gradient	7.7228	8.9130	3.7768	12.3834	0.6236 0.7198 0.3050

	Step				
	x_1	x_2	x_3	Length	x_1 x_2 x_3
<i>DFP - Iteration 1</i>					
Newton-Raphson	6.7466	7.6270	4.8255	11.2682	0.5987 0.6769 0.4282
Conjugate Gradient	4.1395	5.0091	1.7613	6.7327	0.6148 0.7440 0.2616
<i>DFP - Iteration 2</i>					
Newton-Raphson	6.7466	7.6270	4.8255	11.2682	0.5987 0.6769 0.4282
Conjugate Gradient	4.1395	5.0091	1.7613	6.7327	0.6148 0.7440 0.2616
<i>BFGS - Iteration 1</i>					
Newton-Raphson	10.8069	11.9291	5.8015	17.1099	0.6316 0.6972 0.3391
Conjugate Gradient	7.5224	9.0598	3.3966	12.2557	0.6138 0.7392 0.2771
<i>BFGS - Iteration 2</i>					
Newton-Raphson	10.8447	11.7436	6.0193	17.0807	0.6349 0.6875 0.3524
Conjugate Gradient	7.6884	8.9427	3.7078	12.3625	0.6219 0.7234 0.2999

- Iteration 2

$$B_{t+1} = \begin{pmatrix} -0.6315 & 0.4256 & 0.1776 \\ 0.4256 & -0.5083 & 0.2048 \\ 0.0106 & 0.0113 & -0.0846 \end{pmatrix} \quad (3.35)$$

$$B_{t+1}^* = \begin{pmatrix} -0.0628 & 0.0171 & 0.0106 \\ 0.0171 & -0.0612 & 0.0113 \\ 0.0106 & 0.0113 & -0.0846 \end{pmatrix} \quad (3.36)$$

3.3.3.2 Powell Symetric Broyden Update

- Iteration 1

$$B_t = \begin{pmatrix} -0.6703 & 0.3860 & 0.1504 \\ 0.3860 & -0.5565 & 0.1827 \\ 0.1504 & 0.1827 & -0.9365 \end{pmatrix} \quad (3.37)$$

- Iteration 2

$$B_{t+1} = \begin{pmatrix} -0.6328 & 0.4274 & 0.1745 \\ 0.4274 & -0.5109 & 0.2096 \\ 0.1745 & 0.2096 & -0.9223 \end{pmatrix} \quad (3.38)$$

$$B_{t+1}^* = \begin{pmatrix} -0.0629 & 0.0171 & 0.0106 \\ 0.0171 & -0.0612 & 0.0114 \\ 0.0106 & 0.0114 & -0.0850 \end{pmatrix} \quad (3.39)$$

3.3.3.3 Davidon-Fletcher-Powell Update

- Iteration 1

$$B_t = \begin{pmatrix} -0.7187 & 0.4517 & 0.0326 \\ 0.4517 & -0.6455 & 0.3424 \\ 0.0326 & 0.3424 & -1.2232 \end{pmatrix} \quad (3.40)$$

- Iteration 2

$$B_{t+1} = \begin{pmatrix} -0.6788 & 0.4945 & 0.0589 \\ 0.4945 & -0.6021 & 0.3766 \\ 0.0589 & 0.3766 & -0.0971 \end{pmatrix} \quad (3.41)$$

$$B_{t+1}^* = \begin{pmatrix} -0.0653 & 0.0177 & 0.0119 \\ 0.0177 & -0.0602 & 0.0124 \\ 0.0119 & 0.0124 & -0.0971 \end{pmatrix} \quad (3.42)$$

3.3.3.4 Broyden-Fletcher-Goldfarb-Shanno Update

- Iteration 1

$$B_t = \begin{pmatrix} -0.6771 & 0.3952 & 0.1338 \\ 0.3952 & -0.5690 & 0.2051 \\ 0.1338 & 0.2051 & -0.9768 \end{pmatrix} \quad (3.43)$$

- Iteration 2

$$B_{t+1} = \begin{pmatrix} -0.6391 & 0.4369 & 0.1585 \\ 0.4369 & -0.5238 & 0.2333 \\ 0.1585 & 0.2333 & -0.9644 \end{pmatrix} \quad (3.44)$$

$$B_{t+1}^* = \begin{pmatrix} -0.0634 & 0.0172 & 0.0109 \\ 0.0172 & -0.0610 & 0.0115 \\ 0.0109 & 0.0115 & -0.0871 \end{pmatrix} \quad (3.45)$$

3.4 Linear Equality and Inequality Constraints

3.4.1 Linear Equality Constraints

The general optimization problem for the linear equality constraints (LEP) can be stated as

$$\begin{aligned} \min_x & f(x) \\ \text{s.t.} & Ax = b \end{aligned} \quad (3.46)$$

Like the unconstrained problem, we are again looking for the point where the projected gradient vanishes. However, this time instead of searching over dimension n , we only have to search over dimension $n - t$ where t is the number of nonredundant equations in A . In the vernacular of the problem, we want to decompose the vector x into a range-specific portion which is required to solve the constraints and a null-space portion which can be varied. Specifically,

$$x = Yx_Y + Zx_Z \quad (3.47)$$

where Yx_Y denotes the range-specific portion of x and Zx_Z denotes the null-space portion of x . By the definition of the null-space, we know that

$$AYx_Y^* = b \quad (3.48)$$

if x^* is feasible, since $AZ = 0$.

The numerical example given in the text uses a single constraint

$$x_1 + x_2 + x_3 = 3 \quad (3.49)$$

There are several ways to derive the null-space of a matrix. The null-space matrix is

$$Z = \begin{pmatrix} -\frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} \\ \frac{3+\sqrt{3}}{6} & -\frac{3-\sqrt{3}}{6} \\ -\frac{3-\sqrt{3}}{6} & \frac{3+\sqrt{3}}{6} \end{pmatrix} \quad (3.50)$$

```
(gauss) a = 1 1 1 ;
(gauss) null(a);
```

```
-0.57735027 -0.57735027
0.78867513 -0.21132487
-0.21132487 0.78867513
```

Thus, the solution of the constraint can be expressed as

$$x^* = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -\frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} \\ \frac{3+\sqrt{3}}{6} & -\frac{3-\sqrt{3}}{6} \\ -\frac{3-\sqrt{3}}{6} & \frac{3+\sqrt{3}}{6} \end{pmatrix} x_z^* \quad (3.51)$$

where x_z^* is a 2×1 vector. Hence, this maps a change in 2×1 space into a change in 3×1 space maintaining feasibility.

The step for the model algorithm must be selected so that

$$Ap_t = 0 \quad (3.52)$$

Therefore p_t must be a linear combination of the columns in Z

$$Ap_t = 0 \Leftrightarrow p_t = Zp_z \quad (3.53)$$

In this case a $n \times 1$ vector p_z is mapped into p_t by the null-space matrix Z .

3.4.1.1 Algorithm LE

Model algorithm for solving LEP

LE1 *Test for Convergence* Like the UP algorithm the first step is to test whether the current solution is optimal. Again, we rely on the conditions developed in Chapter 2. Specifically, if $|\nabla_x f(x_t) Z| = 0$ the projected gradient is equal to zero and x_t is optimum.

LE2 *Compute a feasible search direction* Compute a nonzero vector p_z , the unrestricted direction of the search. The actual direction of the search is then

$$p_t = Z p_z. \quad (3.54)$$

LE3 *Compute a step length* Compute a positive a_t for which $f(x_t + a_t p_t) < f(x_t)$.

LE4 *Update the estimate of the minimum* $x_{t+1} = x_t + a_t p_t$ and go back to LE1

3.4.1.2 Computation of the Search Direction

As is often the case in this course, the question of the search direction starts with the second order Taylor series expansion. As in the unconstrained case, we derive the approximation of the objective function around some point x_t as

$$f(x_t + p_t) = f(x_t) + \nabla_x f(x_t) p_t + \frac{1}{2} p_t' \nabla_{xx}^2 f(x_t) p_t + \cdots \quad (3.55)$$

Substituting only feasible steps for all possible steps, we derive the same expression in terms of the null-space

$$f(x_t + Z p_z) = f(x_t) + \nabla_x f(x_t) Z p_z + \frac{1}{2} p_z' Z' \nabla_{xx}^2 f(x_t) Z p_z + \cdots \quad (3.56)$$

Solving for the projection based on the Newton-Raphson concept, we derive much the same steps as the constrained optimization problem. The direction for the Steepest Descent algorithm becomes

$$p_t = -Z Z' \nabla_x f(x_t) \quad (3.57)$$

For the Newton Raphson update the search direction becomes

$$p_t = - (Z' \nabla_{xx}^2 f(x_t) Z)^{-1} Z' \nabla_x f(x_t) \quad (3.58)$$

3.4.2 Linear Inequality Constraints

The general optimization problem with linear inequality [LIP] constraints can be written as

$$\begin{aligned} \min_x & f(x) \\ \text{s.t. } & Ax \geq b \end{aligned} \quad (3.59)$$

This problem differs from the linearly constrained problem by the fact that some of the constraints may not be active at a given iteration, or may become active at the next iteration.

3.4.2.1 Algorithm LI

- LI1 *Test for convergence* If the conditions for convergence are met at x_t , terminate. Again the termination condition is that the projected gradient (given the active constraints) has a norm of zero.
- LI2 *Choose which logic to perform* Decide whether to continue minimizing in the current subspace or whether to delete a constraint from the working set. If a constraint is to be deleted go to step LI6. If the same working set is to be retained, go on to step LI3.
- LI3 *Compute a feasible search direction* Compute a vector p_t by applying the null-space equality

$$p_t = Z_t p_z \quad (3.60)$$

- LI4 *Compute a step length* Compute α in this case we must determine if the optimum step length will violate a constraint. Specifically α is equal to the traditional α_t or $\min(\gamma_i)$ which is defined as the minimum distance to a constraint. If the optimum step is less than the minimum distance to another constraint, then go to LI7, otherwise go to LI5.
- LI5 *Add a constraint to the working set* If the optimum step is greater than the minimum distance to another constraint, then you have to add, or make active, the constraint associated with γ_i . After adding this constraint, go to LI7.
- LI6 *Delete a constraint* If the marginal value of one of the Lagrange multipliers is negative, then the associated constraint is binding the objective function suboptimally and the constraint should be eliminated. Delete the constraint from the active set and return to LI1.

LI7 Update the estimate of the solution $x_{t+1} = x_t + \alpha_t p_t$ and go back to LI1.

A significant portion of the discussion in the LIP algorithm centered around the addition or elimination of an active constraint. The concept is identical to the minimum ratio rule in linear programming. Specifically, the minimum ratio rule in linear programming identifies the equation (row) which must leave solution in order to maintain feasibility. The rule is to select that row with the minimum positive ratio of the current right hand side to the a_{ij} coefficient in the matrix. In the nonlinear problem, we define

$$\gamma_i = \left\{ \frac{b_i - a_i' x}{a_i' p} a_i p < 0 \right\} \quad (3.61)$$

3.5 Algorithms for Nonlinear Constraints

The general objective is to maximize or minimize a nonlinear objective function subject to nonlinear constraints. [Gill *et al.* (1981)] lay out the basic problems

$$\begin{aligned} \min_x f(x) \\ \text{s.t. } c_i(x) = 0 \quad i = 1, \dots, t \end{aligned} \quad \text{and} \quad \begin{aligned} \min_x f(x) \\ \text{s.t. } c_i(x) \geq 0 \quad i = 1, \dots, t \end{aligned} \quad (3.62)$$

These optimization problems are much more difficult than the equality or inequality constraints posed in the preceding sections due to the difficulties involved in maintaining feasibility. By the formulation of the null-space in the linear equality scenario, it was always possible to guarantee that x_{t+1} was feasible given that x_t was feasible. Similarly, expansion to linear inequality constraints only added caveats to the step length algorithm and checks on whether a constraint could be deleted. However, the solution of nonlinear constraints may be difficult, if not impossible, without the incorporation of a nonlinear objective function. I want to discuss three different methodologies for optimization under nonlinear constraints. The first two are the use of penalty and barrier functions. The third procedure is the projected augmented Lagrangian method.

3.5.1 Penalty Functions

The penalty and barrier function procedures are very similar. The primary concept of both procedures is to append an additional term onto the ob-

jective function which imposes a cost for violating the constraint. Taking the Penalty function first, assume that we want to optimize

$$\begin{aligned} \min_x x^2 \\ \text{s.t. } x - 1 = 0 \end{aligned} \quad (3.63)$$

Granted that the constraint is linear, and could be solved using a projected quasi-Newton algorithm, a quadratic penalty function for this problem could be specified as

$$P_Q(x, \rho) = x^2 + \frac{\rho}{2}(x - 1)^2 \quad (3.64)$$

where the ρ term yields a penalty for violating the constraint. Solving this problem while increasing the penalty parameter yields an optimal solution with the constraint imposed.

3.5.1.1 Algorithm DP (Model algorithm with a differentiable penalty function)

- DP1 *Check termination conditions.* If x_t satisfies the optimality conditions, terminate with the current solution.
- DP2 *Minimize the penalty function.* Using x_t as the starting point, execute an algorithm to solve the unconstrained subproblem

$$\min_x P_1(x, p_t) \quad (3.65)$$

subject to safeguard conditions to protect against unboundedness. Let x_{t+1} be the estimate of this subproblem.

- DP3 *Increase the penalty parameter.* Set $t + 1$ to a larger value than k . and go back to step DP1.

3.5.2 Barrier Function

The barrier function works well when the constraint can be evaluated either above or below its constrained value. Specifically, if the nonlinear constraint is an equality, then a sufficient penalty parameter can be derived to force that the constraint be close to the right hand side. However, in certain cases, you may want to guarantee feasibility as a minimum condition. For this type of problem the barrier method may be preferred. Focusing on the logarithmic barrier function, we create a subproblem similar to that generated in the penalty parameter framework. Specifically

$$B(x, r) = f(x) - r \sum_{i=1}^m \ln(c_i(x)) \quad (3.66)$$

where r is the barrier parameter and the $\ln(\cdot)$ term in the maximization problem guarantees that the $c_i(x)$ will be greater than zero.

3.5.3 Projected Augmented Lagrangian

Minos 5.1 uses the projected augmented Lagrangian algorithm to optimize problems involving nonlinear constraints. This algorithm as instituted in Minos solves a sequence of subproblems. Each subproblem, or major iteration, solves a linearly constrained minimization (maximization) problem. The constraints for this linear subproblem are the linear constraints plus the linearized nonlinear constraints. Then like the straightforward penalty parameter method, the penalty can be increased to approach the nonlinear constraints with an arbitrary level of precision. Mathematically, the general problem can be written

$$\begin{aligned} \min_x \quad & f(x) + c'x + d'y - \lambda'(G(x) - G(x_t) +) \\ & \frac{1}{2}\rho(G(x) - G(x_t))'(G(x) - G(x_t)) \\ \text{s.t.} \quad & G(x) + A_1y = b_1 \\ & A_2x + A_3y = b_2 \end{aligned} \quad (3.67)$$

$$l \leq \begin{pmatrix} x \\ y \end{pmatrix} \leq u$$

In this problem, the major iteration involves the linearization of the constraints around the point x_t . The first step is to linearize the nonlinear constraints around the starting point of the major iteration

$$\begin{aligned} G(x) &\approx G(x_t) + \nabla_x G(x_t)(x - x_t) \\ &= G(x_t) + \nabla_x G(x_t)x - \nabla_x G(x_t)x_t. \end{aligned} \quad (3.68)$$

Given this relationship, the linear constraints can then be written as

$$\begin{pmatrix} \nabla_x G(x_t) & A_1 \\ A_2 & A_3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 + \nabla_x G(x_t)x_t - G(x_t) \\ b_2 \end{pmatrix} \quad (3.69)$$

Ignoring the slackness conditions.

Chapter 4

Practical Optimization Problems and Computer Code

4.1 Unconstrained Optimization

4.1.1 *Precision Agriculture*

4.1.2 *Estimation of the Inverse Hyperbolic Sine Transformation*

Most courses in statistical inference present the concept of maximum likelihood maximum likelihood. Specifically, given a sample of data points, we can write the likelihood function of a set of parameters as

$$L = \prod_{i=1}^n \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] \quad (4.1)$$

Taking the natural logarithm of this expression yields

$$\ln(L) = \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right]. \quad (4.2)$$

Taking the derivative with respect to μ yields

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \mu} &= \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} = 0 \\ n\mu &= \sum_{i=1}^n x_i \\ \mu &= \frac{1}{n} \sum_{i=1}^n x_i \end{aligned} \quad (4.3)$$

Thus, we derive a maximum likelihood estimate of the mean. In this case the maximum likelihood is "closed form".

Another potential likelihood function involves the inverse hyperbolic sine transformation to normality. This distribution allows variables to be both skewed and kurtotic. [Ramirez *et al.* (1994)] Ramirez O.A. Moss, C. B. Boggess, W. G. estimated corn, soybean and wheat yields over time using this transformation. The log-likelihood function for this distribution with a linear time trend can be written as

$$\ln(L^*) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{i=1}^n \frac{(z_i - \mu)^2}{\sigma^2} - \frac{1}{2} \sum_{i=1}^n (1 + z_i^2 \theta^2) \quad (4.4)$$

$$v_i = y_i - \alpha_0 - \alpha_1 i$$

$$z_i = \frac{\ln \left[v_i \theta + \sqrt{1 + v_i^2 \theta^2} \right]}{\theta}$$

The parameters to be estimated, or maximized with respect to are $\alpha_0, \alpha_1, \theta, \mu$, and σ . Guess what, there exist no closed form solution to those first order conditions.

Several numerical solvers are available for this problem, but I want to outline the solution in GaussTM Aptech. Gauss's optimization routine (for Gauss 3.x) is a subroutine called `optmum`. To use `optmum` you must

- Load the `optmum` library
- Create a subroutine which uses the parameters to be estimated to calculate the value of the objective function (log-likelihood function).
 - You may also want to specify the objective function gradients and the Hessian matrix.
 - Note that `optmum` minimizes the specified objective function. Therefore, you want to specify negative of the likelihood function.
- Initialize the algorithm
 - Initialize by specifying the starting values of the choice variables.
 - Initialize by choosing the an optimization algorithm.
- Call the optimization function.

Example 4.1. Instead of focusing on the more complex example, I want to discuss the estimation of the mean and variance.

Computer Code

61

```
library optmum;
#include optmum.ext;
opset;

load x[20,1]=indta.dta;

proc ml(b);
  local err;
  err=x-b[1];
  retp( rows(err)*ln(b[2])+sumc((x-b[1]).^2 ./ b[2]) );
endp;

b0={10,1.5};
{fhat,bhat,ghat,retcode}=optprt(optmum(&ml,b0));
```

Some notes: First note b is a gauss vector with $b[1]$ containing the value of the mean and $b[2]$ containing the value of the variance. Second, note that the default algorithm is BFGS with numerical gradients. Third, note that estimating the variance instead of the standard deviation raises some potential problems. Why?

Chapter 5

Review of Dynamic Mathematics

5.1 Introduction to Dynamic Optimality

To introduce the problem of dynamic optimization, I would like to start with a discrete time problem. Specifically, we will use the example from Howitt "Some Microeconomics of Agricultural Resource Use." The basic problem is that y_t , the firm's output at time t , is a function of variable inputs, x_t , and the level of capital stock, k_t . Further, capital stock can be augmented by investment, I_t . The firm's overall problem can then be written as

$$\begin{aligned} \max_{x, k, I} \quad & \sum_{t=0}^{T-1} \beta_t \pi(x_t, I_t, k_t) + \beta_T \pi(x_T, k_T) \\ & k_{t+1} - k_t = \delta k_t + I_t \\ & k_0 = \bar{k} \end{aligned} \quad (5.1)$$

Within this formulation, the first equation embodies the objective function, the second equation is referred to as the equation of motion depicting how the capital changes over time, and the third equation is the initial condition on capital.

Rewriting this equation in Lagrangian form yields

$$L = \sum_{t=0}^{T-1} [\beta_t \pi(x_t, I_t, k_t) - \lambda_{t+1} (k_{t+1} - k_t - \delta k_t - I_t)] + \beta_T \pi(x_T, k_T) \quad (5.2)$$

This form of the problem yields three types of conditions. The first type of conditions are with respect to the control variables. This problem has two control variables, x_t and I_t which the decision maker can directly control.

Specifically, the capital variable, k_t , is a state variable which can only be controlled through investment, I_t . The first of the control variables, x_t , has no long term effect. Thus, setting the level of x_t relies on similar rules as a static optimization problem

$$\begin{aligned}\frac{\partial L}{\partial x_t} &= \beta_t \frac{\partial \pi(.)}{\partial x_t} = 0 \\ \frac{\partial \pi(.)}{\partial x_t} &= 0\end{aligned}\tag{5.3}$$

However, the derivative with respect to the investment variable, I_t , yields conditions dependent on λ_{t+1} . λ_{t+1} is the Lagrange multiplier on the equation of motion. As such, it depicts the marginal value of a change in the capital stock. In the jargon of dynamic optimization, this variable is referred to as the costate or adjoint variable. The second condition with respect to I_t is then

$$\frac{\partial L}{\partial I_t} = \beta_t \frac{\partial \pi(.)}{\partial I_t} + \lambda_{t+1} = 0\tag{5.4}$$

Thus, the decision regarding the optimal level of investment must be made considering the future value of the capital stock.

The second condition is the partial with respect to λ_{t+1} , which guarantees that the equation of motion is satisfied. Mathematically

$$\frac{\partial L}{\partial k_t} = \beta_t \frac{\partial \pi(.)}{\partial k_t} - \lambda_t + (1 - \delta) \lambda_{t+1} = 0\tag{5.5}$$

Note that k_t appears in two equations of motion, one at time t and one at time $t + 1$. Taking the derivative of the sum then yields two costate variables in the state equation. This equation can be rewritten as

$$\lambda_{t+1} - \lambda_t = -\beta_t \frac{\partial \pi(.)}{\partial k_t} - \lambda_{t+1} \delta\tag{5.6}$$

Note that this equation can be written as a difference equation

$$\Delta \lambda_{t+1} = -\beta_t \frac{\partial \pi(.)}{\partial k_t} - \lambda_{t+1} \delta\tag{5.7}$$

Thus, the solution of this equation depicts how the costate variable, or the value of the capital constraint, changes over time.

Taking some liberty with the mathematics, and letting Δ_t approach zero. The above equation becomes

$$\lambda'(t) = -\beta(t) \frac{\partial \pi(x(t), I(t), k(t))}{\partial k(t)} - \lambda(t) \delta = 0 \quad (5.8)$$

Which is the continuous representation of the dynamic optimization problem.

This discussion then leads to two different types of dynamic mathematical problems. The first problem, given that time is specified in discrete increments is referred to as a difference equation. The second, given that time is continuous is referred to as a differential equation. In this book, we will emphasize the continuous time formulation.

Finally, the discrete formulation above can be used to demonstrate Pontryagin's optimality conditions. Specifically, the objective function at each increment in time can be written as a function called the Hamiltonian Hamiltonian

$$H_t(x_t, I_t, k_t, \lambda_t) = \beta_t \pi(x_t, I_t, k_t) - \lambda_{t+1}(-\delta k_t - I_t) \quad (5.9)$$

Remembering that the full equation of motion term in the original objective function was

$$\lambda_{t+1}(k_{t+1} - k_t - \delta k_t - I_t) \quad (5.10)$$

The optimal control problem can be rewritten as

$$L = \sum_{t=0}^{T-1} \{H_t(x_t, I_t, k_t, \lambda_t) - \lambda_{t+1}(k_{t+1} - k_t)\} + \beta_T \pi(X_T, k_T) \quad (5.11)$$

This formulation yields the three classic Pontryagin conditions Pontryagin conditions.

(1) The Optimality/Hamiltonian Conditions

$$\frac{\partial H}{\partial u_t} = 0 \Rightarrow \begin{cases} \frac{\partial L}{\partial x_t} = \beta_t \frac{\partial \pi(\cdot)}{\partial x_t} = 0 \\ \frac{\partial L}{\partial I_t} = \beta_t \frac{\partial \pi(\cdot)}{\partial I_t} + \lambda_{t+1} = 0 \end{cases} \quad (5.12)$$

(2) The Adjoint/Costate Conditions

$$\frac{\partial H}{\partial \lambda_{t+1}} = g(s_t, u_t) = s_{t+1} = s_{t+1} - s_t \Rightarrow -\delta k_t - I_t = k_{t+1} - k_t \quad (5.13)$$

(3) The Equation of Motion Constraint

$$\frac{\partial H}{\partial s_t} = \lambda_{t+1} - \lambda_t \Rightarrow \beta_t \frac{\partial \pi(\cdot)}{\partial k_t} - \lambda_{t+1} \delta = \lambda_{t+1} - \lambda_t \quad (5.14)$$

5.2 Review of Ordinary Differential Equation

Thus, from the discussion of the optimal control framework above, we see that dynamic optimization in continuous time involves solving differential equations. We start by focusing on the mathematics of ordinary differential equations [Tu]

Definition 5.1. Ordinary differential equations (ODE) ordinary differential equation are equations involving total derivatives

$$\phi \left(x, \frac{dx}{dt}, \frac{d^2x}{dt^2}, \dots, \frac{d^nx}{dt^n}, t \right) = 0 \quad (5.15)$$

Definition 5.2. The order of an ODE is given by the highest order derivative in the equation. The degree of an ODE is given by the highest exponent of any derivative.

Example 5.1. the equation

$$\ddot{x} + a\dot{x}^3 + bx + c = 0 \quad (5.16)$$

is a second order ODE of degree 3 where

$$\ddot{x} = \frac{d^2x}{dt^2} \text{ and } \dot{x} = \frac{dx}{dt}. \quad (5.17)$$

An ODE is said to be linear if it is of the first degree in the dependent variable and its derivatives

$$a_0(t) \frac{d^nx}{dt^n} + a_1(t) \frac{d^{n-1}x}{dt^{n-1}} + \dots + a_{n-1}(t) \frac{dx}{dt} + a_n(t) = 0. \quad (5.18)$$

An ODE containing terms of a higher degree (than the first) is called a nonlinear ODE

$$\ddot{x} + a\dot{x} + bx^2 = 0 \quad (5.19)$$

The solution of an ODE is defined as any relationship between variables not involving the derivatives which satisfies the ODE. The solution is also called the integral curve.

Theorem 5.1. *Let $x'(t) = f(x, t)$ be continuous with continuous derivatives $\partial f / \partial x$ for x in the domain D and t in the interval I (i.e., $x \in D$ and $t \in I$) there exists a solution $x^*(t)$ defined uniquely in some neighborhood of (x_0, t_0) satisfying the initial conditions $x^*(t_0) = x_0$.*

5.2.1 First Order Linear Differential Equation

First order linear differential equations are equations of the form

$$a_0(t)\dot{x} + a_1(t)x = G(t) \quad (5.20)$$

The first order linear differential equation can be written as

$$\dot{x} + a(t)x = g(t) \quad (5.21)$$

Given that $a_0(t)$ is not equal to zero. The first solution to be derived is the Homogeneous case where $g(t)$ is set equal to zero

$$\dot{x} + ax = 0 \text{ given } x(0) = x_0 \quad (5.22)$$

To solve this problem, we rewrite the original problem as

$$\begin{aligned} \frac{dx}{dt} &= -ax \\ \frac{dx}{x} &= -adt \\ \int \frac{dx}{x} &= - \int adt \\ \ln(x) &= -at + c \\ x(t) &= Ae^{-at} \end{aligned} \quad (5.23)$$

The second scenario in the first order linear differential equation is the non-homogeneous case. Under this scenario, the right hand side of the differential equation is not zero

$$\dot{x} + ax = b, \quad x(0) = x_0 \quad (5.24)$$

The procedure in [Tu] involves substituting $y = x - b/a$ for x which yields a homogeneous equation

$$y = x - \frac{b}{a} \Rightarrow dy = dx \quad (5.25)$$

Thus the differential equation can be written as

$$\begin{aligned} \dot{y} + a \left(y + \frac{b}{a} \right) &= b \\ \dot{y} + ay &= 0 \\ y(t) &= y_0 e^{-at} \\ x(t) &= \left(x_0 - \frac{b}{a} \right) e^{at} + \frac{b}{a} \end{aligned} \quad (5.26)$$

This representation points to the division of the solution into two components. The first component is called the complementary function, which is the solution to the homogeneous part of the problem, $x_c(t) = Ae^{-at}$. The other part is the particular integral which is called the "equilibrium" portion of the solution, $x_p(t) = b/a$. The complete solution is then the sum of those two parts $x(t) = x_c(t) + x_p(t)$.

To develop a fairly general form of the linear problem, let us start by assuming that

$$\dot{x} + ax = g(t). \quad (5.27)$$

Given that $a \neq 0$, this equation actually amounts to finding some function $G(t)$ such that

$$\frac{dx}{dt} + ax = \frac{d}{dt}(G(t)) \quad (5.28)$$

Starting from the solution of the differential equation with constant coefficients, we have $x = c \exp[-at]$. Reversing this expectation, $c = y \exp[at]$. Hence

$$\begin{aligned}\frac{d}{dt}(xe^{at}) &= 0 \\ e^{at}(\dot{x} + ax) &= 0\end{aligned}\tag{5.29}$$

Thus

$$\begin{aligned}\dot{x} + ax &= g(t) \\ e^{at}(\dot{x} + ax) &= e^{at}g(t) \\ d(e^{at}x) &= e^{at}g(t) \\ e^{at}x &= \int_t e^{as}g(s)ds \\ x &= e^{-at} \int_t e^{as}g(s)ds\end{aligned}\tag{5.30}$$

Example 5.2.

$$\begin{aligned}\dot{x} + 3x &= t + e^{-2t} \\ d(xe^{3t}) &= te^{3t} + e^t \\ x &= e^{-3t} \int_t (se^{3s} + e^s) ds + ce^{-3t}\end{aligned}\tag{5.31}$$

Integrating the first term by parts

$$\begin{aligned}\int u dv &= uv - \int v du \\ dv = e^{3s} &\Rightarrow v = \frac{1}{3}e^{3s} \\ u = s &\Rightarrow du = ds \\ \int se^{3s} ds &= \frac{t}{3}e^{3t} - \int \frac{1}{3}e^{3s} ds \\ &= \frac{t}{3}e^{3t} - \frac{1}{9}e^{3t}\end{aligned}\tag{5.32}$$

The solution then becomes

$$x = \frac{t}{3} - \frac{1}{9} + e^{-2t} + ce^{-3t}\tag{5.33}$$

Following the same logic for

$$\dot{x} + p(t)x = g(t)\tag{5.34}$$

we want to choose a function $\mu(t)$ such that

$$\begin{aligned}\mu(t)(\dot{x} + p(t)x) &= (\mu(t)x)' \\ &= \mu'(t)x + \mu(t)\dot{x}\end{aligned}\tag{5.35}$$

which implies that

$$\begin{aligned}d(\mu(t)) &= \mu(t)g(t) \\ \mu(t)x &= \int_t \mu(s)g(s)ds \\ x &= \frac{1}{\mu(t)} \int_t \mu(s)g(s)ds\end{aligned}\tag{5.36}$$

The choice of $\mu(t)$ is then determined by

$$\begin{aligned}\mu(t)(\dot{x} + p(t)x) &= \mu'(t)x + \mu(t)\dot{x} \\ \mu(t)p(t) &= \mu'(t) \\ p(t) &= \frac{\mu'(t)}{\mu(t)} \\ \int_t p(s)ds &= \ln(\mu(t)) \\ \mu(t) &= \exp\left(\int_t p(s)ds\right)\end{aligned}\tag{5.37}$$

5.2.2 Nonlinear Differential Equations

In fairly general terms, we can assert that a first-order linear differential equation has a unique solution. This assertion follows our solution of the

$$\dot{x} + p(t)x = g(t)\tag{5.38}$$

problem. However, solutions for nonlinear differential equations where

$$\dot{x} = x^{\frac{1}{3}} \quad x(0) = 0\tag{5.39}$$

raises some additional difficulties.

Two procedures for solving nonlinear differential equations (1) separable differential equations and (2) exact differential equations. Starting with a

discussion of separable differential equations separable differential equation,
the general expression of a differential equation

$$\frac{dx}{dt} = f(t, x) \quad (5.40)$$

an alternative way to express this relationship is

$$M(t, x) + N(t, x) \frac{dx}{dt} = 0. \quad (5.41)$$

One subgroup of this class of differential equations is referred to as separable differential equations where

$$M(t) + N(x) \frac{dx}{dt} = 0 \quad (5.42)$$

so that each term only involves t or x respectively. Solving this system then involves integrating each section separately

$$\begin{aligned} M(t) &= -N(x) \frac{dx}{dt} \\ M(t) dt &= N(x) dx \\ \int M(t) dt &= \int N(x) dx \end{aligned} \quad (5.43)$$

Example 5.3. Consider the rational function

$$\frac{dx}{dt} = \frac{3t^2 + 4t + 2}{2(x - 1)}, \quad x(0) = -1 \quad (5.44)$$

Separating the function

$$\begin{aligned} 2(x - 1) dx &= (3t^2 + 4t + 2) dt \\ 2 \int_x (z - 1) dz &= \int_t (3s^2 + 4s + 2) ds + c \\ 2 \left(\frac{1}{2} x^2 - x \right) &= (t^3 + 2t^2 + 2t) + c \\ x^2 - 2x &= t^3 + 2t^2 + 2t + c \end{aligned} \quad (5.45)$$

Given that $x(0) = -1$

$$1 - 2(-1) = c \Rightarrow c = 3 \quad (5.46)$$

Thus, the implicit form of the solution is

$$x^2 - 2x = t^3 + 2t^2 + 2t + 3 \quad (5.47)$$

Example 5.4. Next consider

$$\frac{dx}{dt} = \frac{t^3}{x}. \quad (5.48)$$

Separating the equation

$$\begin{aligned} x dx &= t^2 dt \\ \int_x z dz &= \int_t s^2 ds + c \\ \frac{1}{2}x^2 &= \frac{t^3}{3} + c \\ x^2 &= \frac{2}{3}t^3 + c \end{aligned} \quad (5.49)$$

The exact differential equation exact differential equation technique starts from a general equation

$$\psi(t, x) = c \quad (5.50)$$

taking the total differential of this equation with $dc = 0$ we get

$$\psi_t(t, x) + \psi_x(t, x) dx = 0 \text{ or } \psi_t(t, x) + \psi_x(t, x) \frac{dx}{dt} = 0. \quad (5.51)$$

Obviously, we see that since

$$\begin{aligned} \psi_t(t, x) + \psi_x(t, x) \frac{dx}{dt} &= 0 \\ \psi(t, x) &= \int_t \psi_s(s, x) ds \text{ or } \psi = \int_x \psi_z(t, z) dz \end{aligned} \quad (5.52)$$

Spotting this symmetry between the terms allows for the solution of these problems.

Example 5.5. Consider

$$2tx^3 + 3t^2x^2 \frac{dx}{dt} = 0 \quad (5.53)$$

which implies that

$$\psi(t, x) = t^2x^3 \text{ or } t^2x^3 = c \Rightarrow x(t) = t^{-\frac{2}{3}}c \quad (5.54)$$

In general this form can be spotted by Young's theorem. Specifically,

$$\frac{d}{dt} \left(\frac{d}{dx} \psi(t, x) \right) = \frac{d}{dt} \left(\frac{d}{dx} \psi(t, x) \right) \quad (5.55)$$

In Example 5.5

$$\frac{d}{dt} \psi(t, x) = \psi_t(t, x) 2tx^3 \text{ and } \frac{d}{dx} \psi_t = 6tx^2 \quad (5.56)$$

To confirm the exact nature of the solution

$$\frac{d}{dx} \psi(t, x) = \psi_x(t, x) = 3t^2 x^2 \text{ thus } \frac{d}{dx} \psi_x(t, x) = 6tx^2. \quad (5.57)$$

5.2.3 Second and Higher Order Linear Differential Equations

The general n^{th} degree differential equation can be written as

$$a_0(t) x^{(n)} + a_1(t) x^{(n-1)} + \cdots a_{n-1}(t) \dot{x} + a_0(t) x = g(t) \quad (5.58)$$

where $x^{(n)} = d^n x / dt^n$. Introducing the general differential operator $D^i = d^i / dt^i$

$$\begin{aligned} D^2 x(t) &= \frac{d^2 x}{dt^2} = \ddot{x} \\ D^1 x(t) &= \frac{dx}{dt} = \dot{x} \\ D^0 x(t) &= x \end{aligned} \quad (5.59)$$

The n^{th} degree linear differential equation can be written as

$$\begin{aligned} [a_0(t) D^n + a_1(t) D^{n-1} + \cdots a_{n-1}(t) D^1 + a_n(t)] x(t) &= g(t) \\ L(D) x(t) &= g(t) \end{aligned} \quad (5.60)$$

We start by solving for the complementary solution complementary solution. Following the first-order solution, we would posit that the solution of an n^{th} order linear differential equation would be some variant of $x(t) = \exp(\lambda t)$. Taking a second order differential equation

$$\begin{aligned}
[D^2 + bD + c]x &= \ddot{x} + b\dot{x} + c = 0 \\
D^2x &= \lambda^2 e^{\lambda t} \\
Dx &= \lambda e^{\lambda t} \\
x &= e^{\lambda t}
\end{aligned} \tag{5.61}$$

Substituting these values back into the differential equation yields

$$e^{\lambda t} (\lambda^2 + b\lambda + c) = 0 \tag{5.62}$$

Because $\exp(\lambda) \neq 0$

$$\lambda^2 + b\lambda + c = 0 \Rightarrow \lambda = \frac{-b \pm \sqrt{b^2 - 4c}}{2} \tag{5.63}$$

Example 5.6. Starting with constant $a_i(t)$ s

$$\ddot{x} + 2\dot{x} - 3x = 0 \tag{5.64}$$

The characteristic equation for this problem becomes

$$\begin{aligned}
\lambda^2 + 2\lambda - 3 &= 0 \\
\lambda &= \frac{-2 \pm \sqrt{4 + 12}}{2} = \frac{-2 \pm \sqrt{16}}{2} = 1, -3
\end{aligned} \tag{5.65}$$

Thus, there are two possible solutions to this problem

$$\begin{aligned}
x(t) = e^t &\Rightarrow \dot{x} = e^t, \ddot{x} = e^t \\
\ddot{x} + 2\dot{x} - 3x &= e^t + 2e^t - 3e^t = 0 \\
&\text{or} \\
x(t) = e^{-3t} &\Rightarrow \dot{x} = -3e^{-3t}, \ddot{x} = 9e^{-3t} \\
\ddot{x} + 2\dot{x} - 3x &= 9e^{-3t} - 6e^{-3t} - 3e^{-3t} = 0
\end{aligned} \tag{5.66}$$

Thus, this problem has two solutions. Among other things, we are interested in whether these solutions are independent. The independence of the solutions is determined by the Wronskin

$$\begin{aligned}
W(t) &= \begin{vmatrix} x_1(t) & x_2(t) \\ \dot{x}_1(t) & \dot{x}_2(t) \end{vmatrix} \\
&= \begin{vmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ \lambda_1 e^{\lambda_1 t} & \lambda_2 e^{\lambda_2 t} \end{vmatrix} \\
&= e^{-\lambda_1 t} \lambda_2 e^{\lambda_2 t} - e^{\lambda_2 t} \lambda_1 e^{\lambda_1 t} \\
&= e^{(\lambda_1 + \lambda_2)t} (\lambda_2 - \lambda_1) \\
&= e^{(\lambda_1 + \lambda_2)t} \begin{vmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{vmatrix}
\end{aligned} \tag{5.67}$$

In the Example 5.6

$$W(t) \propto \begin{vmatrix} 1 & 1 \\ 1 & -3 \end{vmatrix} = -3 - 1 = -4 \neq 0 \tag{5.68}$$

so the solutions are independent

$$x(t) = c_1 e^t + c_2 e^{-3t} \tag{5.69}$$

5.2.4 Method of Undetermined Coefficients

The method of undetermined coefficients is best described as the method of good guesses. The simplest way to describe this method is by demonstration. Assume that you have the differential equation

$$\ddot{x} + \dot{x} - 2x = 2t. \tag{5.70}$$

A good guess of the particular solution would be

$$\begin{aligned}
x_P(t) &= at + b \\
\dot{x}_P(t) &= a \\
\ddot{x}_P(t) &= 0 \\
\Rightarrow \ddot{x} + \dot{x} - 2x &= 0 + a - 2(at + b) = 2t \\
-2at + (a - 2b) &= 2t.
\end{aligned} \tag{5.71}$$

This solution implies

$$\begin{aligned}
-2at = 2t &\Rightarrow a = -1 \\
a - 2b = 0 &\Rightarrow b = -\frac{1}{2}
\end{aligned} \tag{5.72}$$

Checking the solution

$$\begin{aligned}x_P(t) &= -t - \frac{1}{2} \\ \dot{x}_P(t) &= -1 \\ \ddot{x}_P(t) &= 0 \\ \ddot{x} + \dot{x} - 2x &= 0 = 1 - 2\left(-t - \frac{1}{2}\right) = 2t\end{aligned}\tag{5.73}$$

Example 5.7. Consider the differential equation

$$2\ddot{x} - 4\dot{x} - 6x = 3e^{2t}.\tag{5.74}$$

As an initial guess, we assume that the solution will follow $x_P(t) = Ae^{2t}$ implying

$$\begin{aligned}\dot{x}_P(t) &= 2Ae^{2t} \\ \ddot{x}_P(t) &= 4Ae^{2t} \\ \Rightarrow 2(4Ae^{2t}) - 4(2Ae^{2t}) - 6(Ae^{2t}) &= 3e^{2t} \\ \Rightarrow 8A - 8A + 6A &= 3 \\ \Rightarrow A &= -\frac{1}{2}\end{aligned}\tag{5.75}$$

5.2.5 Laplace Transformed

Laplace transformation

5.2.5.1 Defining Laplace Transformations

One tool in the solution of linear differential equations is the concept of an integral transform. An integral transform is a relationship of the form

$$F(s) = \int_{\alpha}^{\beta} K(s, t) f(t) dt\tag{5.76}$$

where the function $f(t)$ is transformed into another function $F(s)$ by means of the integration. The function F is the transform of f with K being referred to as the kernel of the transformation. The general idea is to transform f into a simpler problem F which can be more easily solved. Afterwards, the original solution f can be recovered. The problem is then to choose an appropriate kernel and bounds of integration.

One possibility for the selection of the kernel was proposed by Laplace. The Laplace transpose uses $K(s, t) = \exp(-st)$, yielding

$$L(f(t)) = F(s) = \int_0^\infty e^{-st} dt. \quad (5.77)$$

This solution is particularly associated with linear differential equations with constant coefficients. Since the Laplace transformation is defined as the proper integral from zero to infinity, it is often useful to introduce the limiting concept

$$\int_a^\infty f(t) dt = \lim_{A \rightarrow \infty} \int_a^A f(t) dt \quad (5.78)$$

Example 5.8. Consider

$$\begin{aligned} L(1) &= \int_0^\infty e^{-st} dt = \lim_{A \rightarrow \infty} \int_0^A e^{-st} dt \\ &= \lim_{A \rightarrow \infty} \left(-\frac{1}{s} e^{-st} \Big|_0^A \right) \\ &= \lim_{A \rightarrow \infty} \left(-\frac{1}{s} e^{-sA} + \frac{1}{s} \right) \\ &= \frac{1}{s} \end{aligned} \quad (5.79)$$

Example 5.9. Consider $f(t) = \exp(at)$

$$\begin{aligned} L(e^{at}) &= \int_0^\infty e^{-st} e^{at} dt = \int_0^\infty e^{-(s-a)t} dt \\ &= \left(\frac{1}{s-a} e^{-(s-a)t} \Big|_0^\infty \right) \\ &= \frac{1}{s-a} \end{aligned} \quad (5.80)$$

In order to use the Laplace transform, we note (without proof) that

$$L[f'(t)] = sL[f(t)] - f(0) \quad (5.81)$$

Similarly, the Laplace transform of a second order differential equation is

$$L[f''(t)] = s^2 L[f(t)] - f(0) \quad (5.82)$$

Thus, the second order differential equation $\ddot{y} - \dot{y} - 2y = 0$ can be written as

$$\begin{aligned} L[y''] - L[y'] - 2L[y] &= 0 \\ [s^2 L[y] - sy(0) - y'(0)] - [sL[y] - y(0)] - 2L[y] &= 0 \end{aligned} \quad (5.83)$$

Given the initial conditions $y(0) = 1$ and $y'(0) = 0$, this expression can be written as

$$\begin{aligned} (s^2 - s - 2)Y(s) + (1 - s)y(0) - y'(0) &= 0 \\ Y(s) &= \frac{s - 1}{(s - 2)(s + 1)} \end{aligned} \quad (5.84)$$

Solving this problem, then involves finding the function whose Laplace transform can be expressed as $Y(s)$.

The Laplace transform of a general function. Taking the Laplace transform of the differential equation above is the same as taking the Laplace transform of each individual term by the additive nature of integration. Thus, to demonstrate the general form of the Laplace transformation, let's take the Laplace transform of a general function $f'(x)$ or the second term in the general differential equation, integrating by parts yields

$$\begin{aligned} \int_0^A f'(t) e^{-st} dt &= f(t) e^{-st} \Big|_0^A + s \int_0^A f(t) e^{-st} dt \\ &= f(A) e^{-sA} - f(0) e^0 + sL[f(t)] \\ &= sL[f(t)] - f(0) \end{aligned} \quad (5.85)$$

Assuming that $f(A) e^{-sA}$ approaches zero as A approaches infinity.

By extension, the Laplace transform of the second differential is

$$\begin{aligned} \int_0^A f''(t) e^{-st} dt &= f'(t) e^{-st} \Big|_0^A + s \int_0^A f'(t) e^{-st} dt \\ &= f'(t) e^{-st} \Big|_0^A + sf(t) e^{-st} \Big|_0^A + s^2 \int_0^A f(t) e^{-st} dt \\ &= -f'(0) - sf(0) + s^2 L[f(t)] \end{aligned} \quad (5.86)$$

with the second expression in Equation 5.86 by integrating the the last term in the first expression by parts.

Given the Laplace transform of the differential equation given above, to solve the differential equation we want to find the function whose Laplace transform matches the derived Laplace transform. To simplify the derivation in this case we want to use the method of partial fractions.

The method of partial fractions involves rewriting the equation

$$L[y] = \frac{s-1}{(s-1)(s+1)} \text{ as } L[y] = \frac{A}{(s-2)} + \frac{A}{(s+1)} \quad (5.87)$$

Much like the method of undetermined coefficients, this is accomplished by assuming that A and B exist, in which case the combined denominator can be written as

$$\begin{aligned} L[y] &= \frac{As + A + Bs - 2B}{(s-2)(s+1)} \\ &= \frac{(A+B)s - (A-2B)}{(s-2)(s+1)} \end{aligned} \quad (5.88)$$

Equating this result with the original fraction yields two equations

$$\begin{aligned} A + B &= 1 \Rightarrow A = 1 - B \\ A - 2B &= -1 \Rightarrow (1 - B) - 2B = -1 \\ 1 - 3B &= -1 \\ B &= \frac{2}{3} \\ A &= \frac{1}{3} \end{aligned} \quad (5.89)$$

Thus, the original equation can be written as

$$L[y] = \frac{1/3}{s-2} + \frac{2/3}{s+1} \quad (5.90)$$

Given this expression, it is possible to find a function form with the same Laplace transform. Specifically, the Laplace transform for $f(x) = \exp(at)$ is

$$\begin{aligned} \int_0^\infty e^{at} e^{-st} dt &= \int_0^\infty e^{-s(s-a)t} dt \\ &= \frac{1}{s-a} \end{aligned} \quad (5.91)$$

Thus, for the first fraction we see that $a = 2$ and for the second fraction we see that $a = -1$. The solution of the differential equation is then

$$y(x) = \frac{1}{3}e^{2x} + \frac{2}{3}e^{-x} \quad (5.92)$$

Example 5.10. Starting with $y'' - y' - 6y = 0$, $y(0) = 1$, and $y'(0) = -1$. The Laplace transformation of this differential equation is

$$\begin{aligned} [s^2 L[y] - sy(0) - y'(0)] - [sL[y] - y(0)] - 6L[y] &= 0 \\ [s^2 L[y] - s + 1] - [sL[y] - 1] - 6L[y] &= 0 \\ [s^2 - s - 6] L[y] &= s - 2 \\ L[y] &= \frac{s - 1}{(s + 2)(s - 3)} \end{aligned} \quad (5.93)$$

Again using partial fractions we have

$$\begin{aligned} L[y] &= \frac{A}{(s + 2)} + \frac{B}{(s - 3)} \\ &= \frac{As - 3A + Bs + 2B}{(s + 2)(s - 3)} \\ A + B &= 1 \Rightarrow A = B - 1 \\ 3A + 2B &= -2 \Rightarrow 3(B - 1) + 2B = -2 \\ 5B &= 1 \text{ or } B = 1/5 \text{ and } A = 4/5 \end{aligned} \quad (5.94)$$

Thus, the separated Laplace transform becomes

$$L[y] = \frac{4/5}{s + 2} + \frac{1/5}{s - 3} \quad (5.95)$$

The solution of the differential equation is then

$$y(x) = \frac{4}{5}e^{2x} + \frac{1}{5}e^{3x} \quad (5.96)$$

5.3 Difference Equations

The discrete time analog of the differential equation is the difference equation. In general, the difference equation grows out of a discrete time representation of some time dependent process

$$y_{t+1} + b(t)y_t = g(t) \quad (5.97)$$

Letting $b(t) = -1$, this expression becomes

$$\begin{aligned} y_{t+1} - y_t &= g(t) \\ \Delta y_t &= g(t) \end{aligned} \quad (5.98)$$

The later form is referred to as a difference equation. Extending this concept,

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) \\ &= \Delta(y_t + 1 - y_t) \\ &= y_{t+2} - 2y_{t+1} + y_t\end{aligned}\tag{5.99}$$

Given this development, we can draw a rough equivalence between Δ^i notation in difference equations and the D^i notation in differential equations.

Definition 5.3. Ordinary difference equations are equations involving one variable x_t measured discretely at different times.

Definition 5.4. The solution of a difference equation is all the values of x_t not involving differences that satisfy the initial difference equation.

Like the differential equation, the solution of a difference equation includes a complementary solution and a particular or equilibrium solution.

5.3.1 First Order Linear Difference Equations

Because of space constraints, we will focus on linear difference equations of the first and second order. Like differential equations, difference equations fall into homogeneous and non-homogeneous. A first order linear homogeneous difference equation can be written as

$$x_{t+1} = ax_t\tag{5.100}$$

Starting with a point $x(0) = x_0$, it is possible to infer the solution of this difference equation

$$\begin{aligned}x_1 &= ax_0 \\ x_2 &= ax_1 = a(ax_0) = a^2x_0 \\ x_3 &= ax_2 = a(a^2x_0) = a^3x_0 \\ x_t &= a^t x_0\end{aligned}\tag{5.101}$$

A first order linear non-homogeneous difference equation can be written as

$$x_t + 1 = ax_t + b\tag{5.102}$$

Again by iteration the solution of this difference equation can be shown to be

$$\begin{aligned}
 a_{t+1} &= ax_t + b \\
 a_{t+2} &= a^2x_t + (1+a)b \\
 a_{t+3} &= a^3x_t + (1+a+a^2)b \\
 a_{t+n} &= a^nx_t + (1+a+a^2+\dots+a^{n-1})b
 \end{aligned} \tag{5.103}$$

In general, the term $(1+a+a^2+\dots+a^{n-1})$ can be rewritten as $1/(1-a)$

$$\begin{aligned}
 \frac{1}{1-a} &\Rightarrow 1 - 1 \times (1-a) = +a \\
 &\Rightarrow a - a \times (1-a) = +a^2 \\
 &\Rightarrow a^2 - a^2 \times (1-a) = +a^3 \\
 &\vdots
 \end{aligned} \tag{5.104}$$

The sum to the n^{th} degree is then derived as

$$\begin{aligned}
 \frac{1-a^n}{1-a} &= \frac{1}{1-a} - \frac{a^n}{1-a} \\
 &= (1+a+a^2+\dots+a^{n-1}+a^n+a^{n+1}+\dots) - (a^n+a^{n+1}+\dots) \\
 &= (1+a+a^2+\dots+a^{n-1})
 \end{aligned} \tag{5.105}$$

The general solution of the non-homogeneous difference equation can then be written as

$$\begin{aligned}
 x_t &= a^t c + \frac{b}{1-a} \text{ if } a \neq 1 \text{ where } c \equiv x_0 - \frac{b}{1-a}, \text{ or} \\
 x_t &= x_0 + bt \text{ if } a = 1.
 \end{aligned} \tag{5.106}$$

Another way that you see these proofs written is with the lag operator, L . Specifically, the non-homogeneous differential equation can be written as

$$\begin{aligned}
 x_{t+1} &= ax_t + b \\
 x_{t+1} &= aLx_{t+1} + b \\
 (1-aL)x_{t+1} &= b \\
 x_{t+1} &= \frac{b}{(1-aL)} = \frac{b}{(1-a)}
 \end{aligned} \tag{5.107}$$

5.3.2 Second Order Linear Difference Equations

Expanding on the first order linear difference equation, the second order linear difference equation can be written as

$$ax_t + bx_{t-1} + cx_{t-2} = g(t) \quad (5.108)$$

The particular solution given that $g(t)$ is equal to some constant d has three cases. First, if $a + b + c$ is not equal to zero, then the particular solution is the constant k which solves

$$\begin{aligned} ak + bk + ck &= d \\ (a + b + c)k &= d \\ k &= \frac{d}{a + b + c} \end{aligned} \quad (5.109)$$

Second, if $a + b + c$ is equal to zero, try the particular solution that $x_t = kt$

$$\begin{aligned} akt + bk(t-1) + ck(t-2) &= d \\ (a + b + c)kt - (b + 2c)k &= d \\ \text{since } a + b + c &= 0 \\ k &= \frac{-d}{b + 2c} \\ x_P &= -\left(\frac{d}{b + 2c}\right)t \end{aligned} \quad (5.110)$$

Finally, if $a + b + c$ is equal to zero and $b + 2c = 0$, then set $x_t = kt^2$. This results in $x_P = d/(2c)t^2$.

The complementary solution is derived in much the same way as the complementary function for a second-order differential equation. Specifically, we conjecture that the solution $x_c(t) = A\lambda^t$. The second order difference equation can then be rewritten as

$$\begin{aligned} ax_t + bx_{t-1} + cx_{t-2} &= 0 \\ aA\lambda^t + bA\lambda^{t-1} + cA\lambda^{t-2} &= 0 \\ A\lambda^{t-2}(a\lambda^2 + b\lambda + c) &= 0 \end{aligned} \quad (5.111)$$

Since $A\lambda^{t-2}$ cannot be zero

$$\begin{aligned} a\lambda^2 + b\lambda + c &= 0 \\ \lambda &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned} \quad (5.112)$$

Just like the solution to the second order differential equation, this solution yields three possibilities. (1) Both roots are real and distinct. In which case the solution to the difference equation is

$$x_t = A_1 \lambda_1^t + A_2 \lambda_2^t. \quad (5.113)$$

(2) If both roots are real, but identical then

$$x_t = (A_1 + A_2 t) \lambda^t. \quad (5.114)$$

(3) If $b^2 - 4ac$ is negative, then the solutions are imaginary with the same style solution as found in the differential equation.

Chapter 6

Dynamic Optimality

6.1 Overview of Dynamic Optimization

The basic calculus problem is to choose the levels of variable x in a way that

$$\max_x F(x). \quad (6.1)$$

We are interested in expanding this problem into the time dimension. One way of incorporating the time dimension is simply

$$\begin{aligned} \max_{x(t)} \int_0^T F(t, x(t)) dt \\ \text{s.t. } x(t) \geq 0 \end{aligned} \quad (6.2)$$

However, this problem is separable in that it can be broken down into several separate problems. Remember x_t in the discrete optimization problem. Another specification that is truly dynamic is

$$\begin{aligned} \max_{x(t), u(t)} \int_0^T F(t, x(t), u(t)) dt \\ \text{s.t. } x'(t) = g(t, x(t), u(t)) \end{aligned} \quad (6.3)$$

In this approach we are trying to derive the path or a function $u(t)$ that describes the control process which adjusts $x(t)$ in a way so as to maximize "profit". The resulting conditions are

$$\begin{aligned}
\frac{\partial H}{\partial u} &= 0 \Rightarrow f_u + \lambda g_u = 0 \\
-\frac{\partial H}{\partial x} &= \lambda'(t) \Rightarrow \lambda'(t) = -(f_x + \lambda g_x) \cdot \\
\frac{\partial H}{\partial \lambda} &= x'(t) \Rightarrow x'(t) = g(t, x(t), u(t))
\end{aligned} \tag{6.4}$$

To arrive at this rule, we will first examine the calculus of variations formulation

$$\begin{aligned}
&\max_{x(t)} \int_0^T F(t, x(t), x'(t)) dt \\
&\text{s.t. } x(t) \geq 0, x(0) = x_0
\end{aligned} \tag{6.5}$$

6.2 Simplified Example

Suppose a firm has received an order for B units of product to be delivered by time T . It seeks a production function schedule for filling this order at the specified delivery data at minimum cost, bearing in mind that unit production cost rises linearly with the production rate and that the unit cost of holding inventory per unit time is constant. Let $x(t)$ denote the inventory accumulated by time t . Then we have $x(0) = 0$ and must achieve $x(T) = B$. The inventory level, at any moment, is the cumulated past production; the rate of change of inventory is the production rate $dx/dt = x'(t)$. Thus, the firm's total cost at any moment (t) is

$$c_1 x'(t) x'(t) + c_2 x(t) = c_1 [x'(t)]^2 + c_2 x(t). \tag{6.6}$$

The cost of producing rises linearly in production (i.e., the rate of change in the level of inventory) the $pc = c_1 x'(t)$ as depicted in Figure 6.1. Thus, the total cost of production is $pc x'(t) = c_1 [x'(t)]^2$. The second term in Equation 6.6 is unit cost of holding inventory.

The firm's objective is to determine the production plan $x'(t)$ that minimizes the total cost to the firm. Thus

$$\begin{aligned}
&\min \int_0^T [c_1 (x'(t))^2 + c_2 x(t)] dt \\
&\text{s.t. } x(0) = 0, x(T) = B, x'(t) \geq 0
\end{aligned} \tag{6.7}$$

One possible plan is to produce at a uniform rate $x'(t) = B/T$. Then,

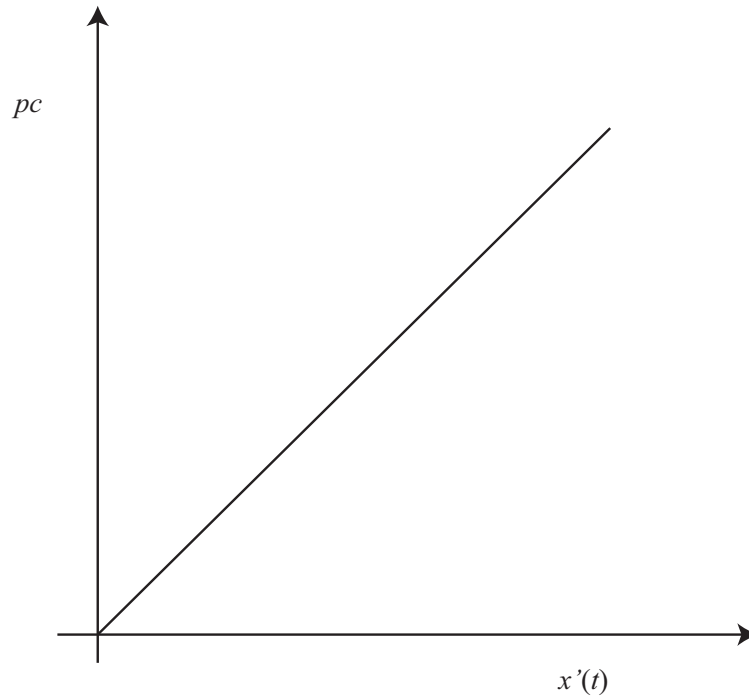


Fig. 6.1 Linear Cost Function

$$x(t) = \int_0^T \left(\frac{B}{T} \right) dt = \frac{BT}{T} \quad (6.8)$$

and the cost incurred will be

$$\int_0^T \left[c_1 \left(\frac{B}{T} \right)^2 + c_2 \frac{Bt}{T} \right] dt = c_1 \frac{B^2}{T} + c_2 \frac{BT}{2} \quad (6.9)$$

while this plan is feasible, it may not minimize costs.

Assume $c_2 = 0$ so the problem becomes

$$\begin{aligned} \min \int_0^T [x'(t)] dt \\ \text{s.t. } x(0) = 0, x(T) = Bx'(t) \geq 0 \end{aligned} \quad (6.10)$$

The solution mechanism presented here is nonstandard. It involves chopping the continuous function into discrete intervals and maximizing. The

question then arises: Why attempt this nonstandard solution? Why not just begin with other methods? First, like the case of the discrete model presented in Chapter 5, the discrete model here has an intuitive appeal. Second, some more advanced applications of optimal control are based on a similar decomposition.

Divide the interval $[0, T]$ into T/k segments of equal length k . Hence, $x(t)$ can be approximated by a piecewise continuous function at points $(0, 0)$, (k, y_1) , $(2k, y_2)$, $\dots$, (T, B) as depicted in Figure ?? . Hence, $x'(t)$ can be approximated by

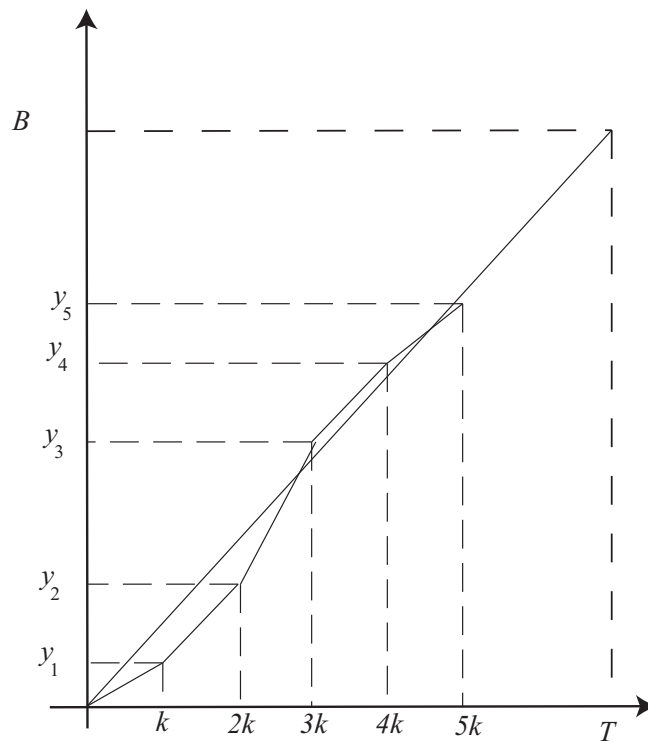


Fig. 6.2 Discrete Control Path

$$\frac{\Delta x}{\Delta t} = \frac{(y_i - y_{i-1})}{k}. \quad (6.11)$$

Thus, the approximating problem is to find y_i $i = 1, \dots, (T/k) - 1$ to

$$\min \sum_{i=1}^{T/k} \left[\frac{(y_i - y_{i-1})}{k} \right]^2 \quad (6.12)$$

(Note that k approximates dt) where $y_0 = 0$ and yT/k . Taking the derivative with respect to y_i yields

$$\frac{\partial}{\partial y_i} \sum_{i=1}^{T/k} \left[\frac{(y_i - y_{i-1})}{k} \right]^2 = \frac{2(y_i - y_{i-1})}{k} - \frac{2(y_{i+1} - y_i)}{k} = 0. \quad (6.13)$$

$$y_i - y_{i-1} = y_{i+1} - y_i$$

The change in y_i s are equal for each period. Thus, the inventory grows linearly throughout the period. The continuous time equivalent is then found by letting $k \rightarrow 0$. The preceding calculation then implies that

$$x(t) = \frac{B}{T}t \quad (6.14)$$

so that

$$x'(t) = \frac{B}{t} \geq 0. \quad (6.15)$$

Here is the tie in bit. The conjecture that the constant path is the least cost path can be verified in this case by showing that no other feasible path exists with a lower cost. First note that all feasible paths must obey the terminal conditions. Hence, let us define $z(t)$ as a continuously differentiable function satisfying $z(0) = 0$ and $z(T) = B$. $z(t)$ is, thus, a feasible path. Define $h(t) = z(t) - x(t)$. Thus, $h(t)$ is defined as the difference between the new feasible path and the old feasible path

$$h(t) = z(t) - \frac{Bt}{T}. \quad (6.16)$$

Note that $h(0) = h(T) = 0$ for a feasible path. Thus,

$$\begin{aligned} z(t) &= h(t) + \frac{Bt}{T} \\ z'(t) &= h'(t) + \frac{B}{T} \end{aligned} \quad (6.17)$$

The difference in cost is then

$$\begin{aligned}
\int_0^T \left\{ [z'(t)]^2 - [x'(t)]^2 \right\} dt &= \int_0^T \left\{ \left[\frac{B}{T} + h'(t) \right]^2 - \left[\frac{B}{T} \right]^2 \right\} dt \\
&= \int_0^T \left\{ \left(\frac{B}{T} \right)^2 + \frac{2B}{T} h'(t) + [h'(t)]^2 - \left(\frac{B}{T} \right)^2 \right\} dt \\
&= \frac{2B}{T} \int_0^T h'(t) dt + \int_0^T [h'(t)]^2 dt \\
&= \frac{2B}{T} (h(T) - h(0)) + \int_0^T [h'(t)]^2 dt \geq 0
\end{aligned} \tag{6.18}$$

Thus, any other production plan will give a higher cost of production.

6.3 Euler Equation

In general, we are trying to find the properties of the function $x(t)$ that solves

$$\begin{aligned}
&\max \int_{t_0}^{t_1} F(t, x(t), x'(t)) dt \\
&\text{s.t. } x(t_0) = x_0, x(t_1) = x_1
\end{aligned} \tag{6.19}$$

6.3.1 A Mathematical Digression

Several points in the derivation of the calculus of variations have strong parallels to derivations in basic calculus. Hence some review would be prudent.

6.3.1.1 The Calculus of Variations - A Background

First, we are familiar with the straightforward definition of a function

$$y = f(x). \tag{6.20}$$

Thus, y is a magnitude that depends on or is a function of x . The calculus of variations is based on a reorientation of this terminology

$$A = \int_a^b f(x) dx. \tag{6.21}$$

A is now a magnitude that is a function of a function. Hence, a functional is a function of a function. The point of the example is only to illustrate that we are looking for a function of x so as to maximize a *functional*, in this case a definite integral of that function; in ordinary calculus we look for the value of a *variable*, x , so as to maximize (or minimize) a known function of that value.

6.3.1.2 Underlying Mathematical Concepts

Maximization: Going back to our definition of a maximum, the point x^* maximizes the value of function $f(x)$ if $f(x) - f(x^*)$ for all x in a neighborhood around x^* . We then developed the traditional first and second order necessary conditions around this point using the Taylor-Series Expansion

$$f(x) - f(a) = f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \cdots \leq 0. \quad (6.22)$$

To apply these conditions, we assumed that the first and second derivatives of the function are continuous.

The approach to the calculus of variations problems will parallel that of ordinary maximization and minimization problems in many respects. In particular, we will need to consider neighborhoods of a function $x(t)$, which is a candidate for the solution to maximize or minimize problems like

$$\max I[x] = \int_{t_0}^{t_1} F(t, x(t), x'(t)) dt. \quad (6.23)$$

Neighborhoods, Continuity, and the Maxima and Minima: In variational calculus, a function $x(x)$ is said to maximize the functional $I[.]$ if $I[y] - I[x] \leq 0$ for all $y(t)$ near to $x(t)$. The use of concepts similar to derivatives and Taylor series for $I[x]$ instead of $f(x)$ requires continuity of the functional; the definition of continuity, in turn, requires a measure of nearness for functions.

Continuity of functional $I[x]$: If for any $\epsilon > 0$ there is a $\delta > 0$ such that $|x_0(t) - x(t)| < \delta$, $|I[x(t)] - I[x_0(t)]| < \epsilon$, then $I[x]$ is continuous at $x_0(t)$.

Paralleling the definition of distance for numbers, the distance between two functions can be defined as the 2-norm of the maximum distance between any two points on the functions measures at the same input value

$$\|x_0(t)\| = \max_{t_0 < t < t_1} |x_0(t)| \quad (6.24)$$

or along the whole function

$$\|x_0(t) - x(t)\| = \max_{t_0 < t < t_1} |x_0(t) - x(t)| \quad (6.25)$$

However, an additional facet beyond traditional calculus is the notion of shape. For example, assume that we are interested in generating a function close to $x(t)$ in Figure 6.3. Using the concept of distance above we could generate a function $x^*(t)$ as depicted in Figure 6.3. Obviously, while $x^*(t)$ lies in with neighborhood defined by δ , its shape deviates significantly from $x(t)$. To circumvent this problem, we extend the notion of distance to include distances for the derivatives of the function

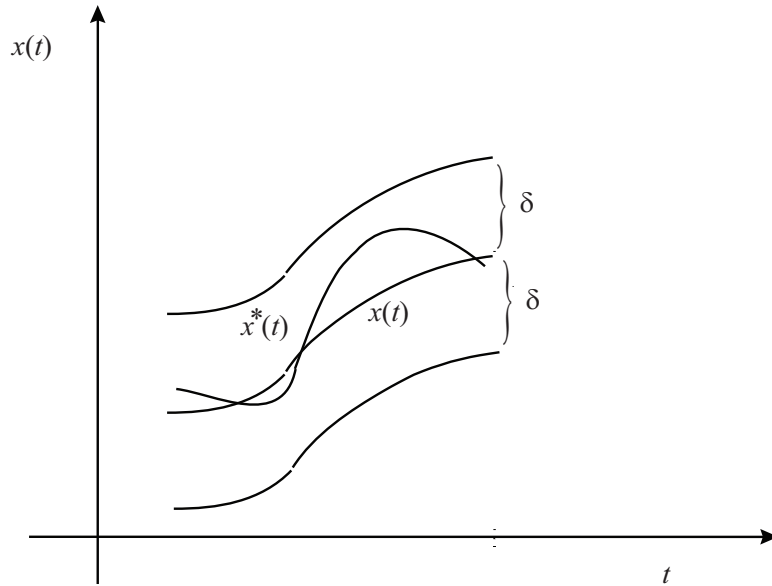


Fig. 6.3 Defining the Neighborhood of a Function

$$\begin{aligned} \|x\|_0 &= \max |x(t)| \\ \|x\|_1 &= \max |x(t)| + \max |x'(t)| \\ \|x\|_2 &= \max |x(t)| + \max |x'(t)| + \max |x''(t)| \end{aligned} \quad (6.26)$$

with similar definitions for the distance between two functions. Using the concept of a neighborhood through the first derivative, we could visualize the neighborhood around any path as depicted in Figure 6.4.

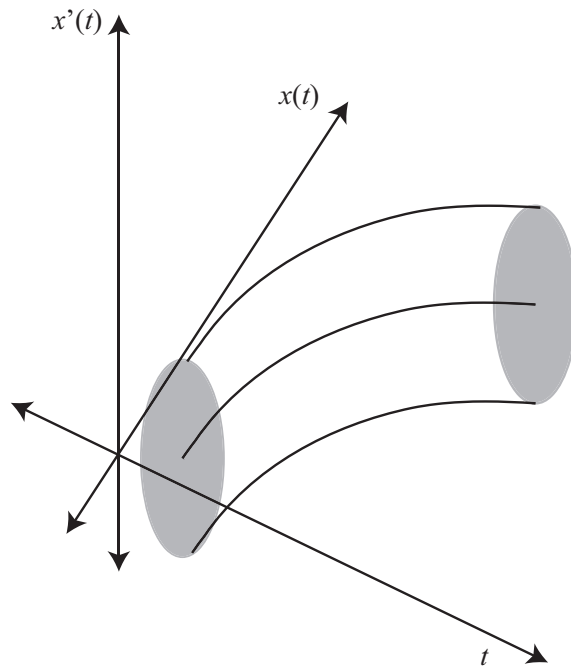


Fig. 6.4 Three Dimensional Neighborhood

6.3.2 Definitions of a Maximum

Definition 6.1 (Weak Maximum). If there is a $\epsilon > 0$ such that for all $\|x_0(t)\| < \epsilon$, $I[x_0(t)] - I[x(t)] \geq 0$, then $x_0(t)$ provides a weak maximum to the functional I .

Definition 6.2 (Strong Maximum). If there is a $\epsilon > 0$ such that for all $\|x_0(t) - x(t)\| < \epsilon$, $I[x_0(t)] - I[x(t)] > 0$, then $x_0(t)$ provides a strong maximum to the functional I .

6.3.3 Necessary Conditions

Assume that $x(t)$ is the optimal curve. Assume that $z(t)$ is in the neighborhood of $x(t)$. One way to think of $z(t)$ is as a perturbation of $x(t)$, or as a *variation* variation of the function $x(t)$ denoted $\delta x(t)$. This is similar to Δx in ordinary calculus; however, the independent variable is often shown explicitly and the notation becomes $\delta x(t)$ which is not δ times $x(t)$. The variation of $x(t)$ can then be written as

$$\begin{aligned} z(t) &= x(t) + \delta x(t) \\ \delta x(t) &= z(t) - x(t) \end{aligned} \quad (6.27)$$

Note that if $z(t)$ is near to $x(t)$ of order 1, $\delta x(t)$ represents an infinitesimal change in $x(t)$ and $\delta x(t)$ represents an infinitesimal change from $x'(t)$. Changing notation again, let

$$\delta x(t) = \epsilon \phi(t) \quad (6.28)$$

where $\phi(t_0) = \phi(t_1) = 0$. To derive the maximum, we will differentiate the functional expressed with respect to the alternative path with respect to ϵ . To be technically correct, we must first define two new functions $\Phi(t)$ and $\Phi'(t)$ based on the $z(t)$ and $z'(t)$ functions. Specifically,

$$\begin{aligned} z(t) &\Rightarrow \Phi(\epsilon) = x(t) + \epsilon \phi(t) \\ z'(t) &\Rightarrow \Phi'(\epsilon) = x'(t) + \epsilon \phi'(t) \end{aligned} \quad (6.29)$$

We will then examine the properties of this system as ϵ approaches zero. Substituting $z(t)$ into the objective yields

$$I = \int_{t_0}^{t_1} F(t, x(t) + \epsilon \phi(t), x'(t) + \epsilon \phi'(t)) dt \quad (6.30)$$

Applying Leibniz's rule

$$\begin{aligned} \frac{d}{d\epsilon} \int_{t_0}^{t_1} F(t, \Phi(t), \Phi'(t)) dt &= \frac{dt_1}{d\epsilon} F(t, \Phi(t), \Phi'(t)) \\ &- \frac{dt_0}{d\epsilon} F(t, \Phi(t), \Phi'(t)) + \int_{t_0}^{t_1} \left[\frac{\partial F}{\partial \Phi} \frac{d\Phi}{d\epsilon} + \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right] dt \end{aligned} \quad (6.31)$$

Under typical assumptions the first and second terms of this differentiation go to zero because the starting point and termination point of the dynamic optimum are exogenous to the optimization problem. However, these conditions can be revisited if the question regards optimum stopping time. More to the point, the Euler equation is derived from the expression under the integral (the last term in the differential). This condition must be met at every time period.

First, we need to carry the integration through the sum yielding

$$\int_{t_0}^{t_1} \left[\frac{\partial F}{\partial \Phi} \frac{d\Phi}{d\epsilon} + \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right] dt = \int_{t_0}^{t_1} \frac{\partial F}{\partial \Phi} \frac{d\Phi}{d\epsilon} dt + \int_{t_0}^{t_1} \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} dt \quad (6.32)$$

Intuitively, we would like to be able to recombine these components in such a way as to factor a common $d\Phi$ term outside the multiplication. To do this we note that the second term in the addition can be integrated by parts

$$\begin{aligned} \int_{t_0}^{t_1} \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} dt &\Rightarrow \begin{cases} dv = \Phi' dt \\ u = \frac{\partial F}{\partial \Phi'} \end{cases} \\ \int_{t_0}^{t_1} \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} dt &= \left(\frac{\partial F}{\partial \Phi'} \frac{d\Phi}{d\epsilon} \right) \Big|_{t=t_1} - \left(\frac{\partial F}{\partial \Phi'} \frac{d\Phi}{d\epsilon} \right) \Big|_{t=t_0} \\ &\quad - \int_{t_0}^{t_1} \left(\frac{d}{dt} \left(\frac{\partial F}{\partial \Phi'} \right) \frac{d\Phi}{d\epsilon} \right) dt \end{aligned} \quad (6.33)$$

Substituting for the derivative of Φ with respect to ϵ this equation becomes

$$\int_{t_0}^{t_1} \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} dt = \frac{\partial F}{\partial \Phi'} \phi(t_1) - \frac{\partial F}{\partial \Phi'} \phi(t_0) - \int_{t_0}^{t_1} \left(\frac{d}{dt} \left(\frac{\partial F}{\partial \Phi'} \right) \frac{d\Phi}{d\epsilon} \right) dt \quad (6.34)$$

Applying the definition of a feasible path the first two terms drop out. Substituting this result back into the original integral and substituting again for the derivative of Φ with respect to ϵ yields

$$\begin{aligned} \int_{t_0}^{t_1} \left[\frac{\partial F}{\partial \Phi} \frac{d\Phi}{d\epsilon} + \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right] dt &= \int_{t_0}^{t_1} \left[\frac{\partial F}{\partial \Phi} \phi(t) - \frac{d}{dt} \left(\frac{\partial F}{\partial \Phi'} \phi(t) \right) \right] dt \\ &= \int_{t_0}^{t_1} \left[\frac{\partial F}{\partial \Phi} - \frac{d}{dt} \left(\frac{\partial F}{\partial \Phi'} \right) \right] \phi(t) dt \end{aligned} \quad (6.35)$$

Finally, letting ϵ approach zero, we derive the Euler equation in terms of the optimal path

$$\frac{dI}{d\epsilon} \Big|_{\epsilon=0} = \int_{t_0}^{t_1} \left[\frac{\partial F}{\partial x} - \frac{d}{dt} \frac{\partial F}{\partial x'}(t) \right] \phi(t) dt = 0 \quad (6.36)$$

implying that

$$f_x - \frac{d}{dt} f_{x'} = 0 \quad \forall t \in [t_0, t_1] \quad (6.37)$$

To operationalize the Euler equation we note that

$$\begin{aligned} f_{x'} &= f_{x'}(t, x(t), x'(t)) \\ \therefore \frac{df_{x'}}{dt} &= f_{x't} + f_{x'x}x' + f_{x'x'}x'' \end{aligned} \quad (6.38)$$

Therefore another manifestation of the Euler equation is

$$f_x = f_{x't} + f_{x'x}x' + f_{x'x'}x'' \quad (6.39)$$

Example 6.1. As a first example, consider solving the simple example presented above

$$\begin{aligned} \min \int_0^T [x'(t)]^2 dt \\ \text{s.t. } x(0) = 0, x(T) = B \end{aligned} \quad (6.40)$$

Solving this example implies that

$$f(t, x(t), x'(t)) = [x'(t)]^2 \Rightarrow \begin{cases} f_x = 0 \\ f_{x'} = 2x'(t) \Rightarrow \begin{cases} f_{x't} = 0 \\ f_{x'x} = 0 \\ f_{x'x'} = 2 \end{cases} \end{cases} \quad (6.41)$$

The Euler equation then becomes

$$f_x = f_{x't} + f_{x'x}x' + f_{x'x'}x'' \Rightarrow 0 = 0 + 0x'(t) + 2x''(t) \quad (6.42)$$

Solving the second order differential equation

$$\begin{aligned} \frac{d^2x(t)}{dt^2} = 0 &\Rightarrow \frac{dx(t)}{dt} = \int 0dt + c_1 = c_1 \\ \frac{dx(t)}{dt} = c_1 &\Rightarrow \int c_1 dt + c_2 = c_1 t + c_2 \end{aligned} \quad (6.43)$$

Using the boundary conditions to solve for the constants

$$\begin{aligned} x(0) = 0 &\Rightarrow 0 = c_1 \cdot 0 + c_2 \Rightarrow c_2 = 0 \\ x(T) = B &\Rightarrow B = c_1 T \Rightarrow c_1 = \frac{B}{T} \end{aligned} \quad (6.44)$$

Thus, the optimal solution is the same as we saw above in Equation 6.18

$$x(t) = \frac{Bt}{T}, \text{ for } 0 \leq t \leq T \quad (6.45)$$

Example 6.2. Expanding the general formulation in Example 6.1

$$\int_0^1 \left\{ [x'(t)]^2 + 10tx(t) \right\} dt \quad (6.46)$$

$$x(0) = 1, \quad x(1) = 2$$

To derive the Euler equation

$$\begin{aligned} f(t, x(t), x'(t)) &= [x'(t)]^2 + 10tx(t) \\ \Rightarrow \begin{cases} f_x = 10t \\ f_{x'} = 2x'(t) \end{cases} &\Rightarrow \begin{cases} f_{x't} = 0 \\ f_{x'x} = 0 \\ f_{x'x'} = 2 \end{cases} \\ &\Rightarrow 10t = 2x''(t) \\ &\Rightarrow x''(t) = 5t \end{aligned} \quad (6.47)$$

The solution of the differential equation is then

$$\begin{aligned} \frac{d^2x(t)}{dt^2} = 5t &\Rightarrow \frac{dx(t)}{dt} = \int 5t dt + c_1 = \frac{5}{2}t^2 + c_1 \\ \frac{dx(t)}{dt} = \frac{5}{2}t^2 + c_1 &\Rightarrow x(t) = \int \left(\frac{5}{2}t^2 + c_1 \right) dt + c_2 = \frac{5}{6}t^3 + c_1t + c_2 \end{aligned} \quad (6.48)$$

Using the initial conditions

$$\begin{aligned} x(0) &= \frac{5}{6}0^3 + c_10 + c_2 = 1 \Rightarrow c_2 = 1 \\ x(1) &= \frac{5}{6}1^3 + c_11 + 1 = 2 \Rightarrow \frac{12}{6} - \frac{5}{6} - \frac{6}{6} = \frac{1}{6} \end{aligned} \quad (6.49)$$

The solution of the calculus of variations problem is then

$$x(t) = \frac{5}{6}t^3 + \frac{1}{6}t + 1 \quad (6.50)$$

Example 6.3. The inventory problem considering not only the cost of production, but also the inventory holding cost then becomes

$$\min \int_0^T \left\{ c_1 [x'(t)]^2 + c_2 x(t) \right\} dt \quad (6.51)$$

$$x(0) = 0 \quad x(T) = B$$

The Euler equation can then be derived as

$$f(t, x, x') = c_1 [x'(t)]^2 + c_2 x(t) \Rightarrow \begin{cases} f_x = 2c_2 \\ f_{x'} = 2c_1 x'(t) \Rightarrow \begin{cases} f_{x't} = 0 \\ f_{x'x} = 0 \\ f_{x'x'} = 2c_1 \end{cases} \\ c_2 = 2c_1 x''(t) \end{cases} \quad (6.52)$$

Solving the Euler equation

$$\begin{aligned} \frac{d^2 x(t)}{dt^2} &= \frac{1}{2} \frac{c_1}{c_2} \Rightarrow \frac{dx(t)}{dt} = \frac{1}{2} \int \frac{c_1}{c_2} dt + k_1 = \frac{c_1}{2c_2} t + k_1 \\ \frac{x(t)}{dt} &= \frac{c_1}{2c_2} t + k_1 \Rightarrow x(t) = \frac{c_1}{2c_2} \int t dt + \int k_1 dt + k_2 \\ &= \frac{c_1}{4c_2} t^2 + k_1 t + k_2 \end{aligned} \quad (6.53)$$

Using the boundary conditions

$$\begin{aligned} x(0) &= \frac{c_1}{4c_2} 0^2 + k_1 0 + k_2 = 0 \Rightarrow k_2 = 0 \\ x(T) &= \frac{c_1}{4c_2} T^2 + k_1 T = B \Rightarrow k_1 = \frac{B}{T} - \frac{c_1}{4c_2} T \\ \therefore x(t) &= \frac{c_1}{4c_2} t^2 + \left(\frac{B}{T} - \frac{c_1}{4c_2} T \right) t = \frac{c_2(t-T)t}{4c_1} + \frac{Bt}{T} \end{aligned} \quad (6.54)$$

In addition to the straightforward solution, we have an implicit condition that $x'(t) \geq 0$. To establish this, we first establish that $x'(0)$ is greater than zero. Afterwards, if we can establish that $x''(t)$ is greater than zero, we will have shown that $x'(t) \geq 0$ for all t in the interval. Starting with the second part first,

$$x''(t) = \frac{c_1}{2c_2} > 0 \text{ if } c_1, c_2 > 0 \quad (6.55)$$

So the result that $x'(t)$ greater than zero is established if $x'(0) \geq 0$

$$\begin{aligned} x'(t) &= \frac{c_1}{2c_2} t + k_1 \\ x'(t) &= \frac{c_1}{2c_2} 0 + \left[\frac{B}{T} + \frac{c_1}{4c_2} T \right] \geq 0 \Leftrightarrow B \geq \frac{c_2 T^2}{4c_1} \end{aligned} \quad (6.56)$$

6.3.4 Interpreting the Euler Equation

First note that the Euler equation must hold between any two points in $0 \leq t \leq T$. Thus, choosing points t and $t + \Delta$ in the previous example yields

$$\begin{aligned} \int_t^{t+\Delta} 2c_1 x''(s) ds &= \int_t^{t+\Delta} c_2 ds \\ 2c_1 \int_t^{t+\Delta} x''(s) ds &= c_2 \int_t^{t+\Delta} ds \\ 2c_1 (x'(t+\Delta) - x'(t)) &= c_2 (t+\Delta - t) \\ 2c_1 x'(t+\Delta) &= 2c_1 x'(t) + c_2 \Delta \end{aligned} \quad (6.57)$$

where $2c_1 x'(t+\Delta)$ is the cost of producing in the next period, $2c_1 x'(t)$ is the cost of producing in the current period, and $c_2 \Delta$ is the inventory cost. Thus, at optimum the cost of producing in the next period is exactly equal to the cost of producing in this period plus inventory cost.

Another form of the Euler condition implicit in the preceding discussion is

$$f_{x'}(t, x(t), x'(t)) = \int_{t_0}^{t_1} f_x(t, x(t), x'(t)) dt \quad (6.58)$$

An example of this reformulation is the inventory problem with a discount rate. This formulation becomes

$$\min \int_0^T e^{-rt} \left\{ c_1 [x'(t)]^2 + c_2 x(t) \right\} dt \quad (6.59)$$

Following the reformulation

$$\begin{aligned} f(t, x(t), x'(t)) &= e^{-rt} \left\{ c_1 [x'(t)]^2 + c_2 x(t) \right\} \\ \Rightarrow \begin{cases} f_x &= e^{-rt} c_2 \\ f_{x'} &= 2e^{-rt} c_1 x'(t) \end{cases} \end{aligned} \quad (6.60)$$

The general statement of the Euler equation then yields

$$f_x = \frac{df_{x'}}{dt}. \quad (6.61)$$

Substituting the results Equation 6.60 into Equation 6.61 yields

$$e^{-rt}c_2 = \frac{d}{dt} [2e^{-rt}c_1x'(t)] \quad (6.62)$$

multiplying both sides of Equation 6.63 by dt , changing the constant of integration to ds and integrating yields

$$\begin{aligned} c_2 \int_t e^{-rs} ds + k_1 &= \int_t d [2e^{-rt}c_1x'(s)] \\ -\frac{c_2}{r}e^{-rt} + k_1 &= 2e^{-rt}c_1x'(t) \\ x(t) &= \int_t \left[-\frac{c_2}{2rc_1} + k_1e^{rs} \right] ds + k_2 \\ &= -\frac{c_2}{2rc_1}t + \frac{k_1}{r}e^{rt} + k_2 \end{aligned} \quad (6.63)$$

Solving for the boundary conditions

$$\begin{aligned} x(0) &= -\frac{c_2}{c_1}0 + \frac{k_1}{r}e^0 + k_2 = 0 \Rightarrow k_2 = -\frac{k_1}{r} \\ x(T) &= -\frac{c_2}{c_1}T + (e^{rt} - 1) \frac{k_1}{r} = B \Rightarrow \frac{r \left[B + \frac{c_2T}{2rc_1} \right]}{[e^{rt} - 1]} \end{aligned} \quad (6.64)$$

Example 6.4 (Maximize Utility over Time). An individual seeks the consumption rate over some time T . The utility function is assumed to obey the typical conditions $U'(x) > 0$, $U''(x) < 0$

$$\max \int_0^T e^{-rt}U(c(t)) dt \quad (6.65)$$

subject to a cash flow constraint. To derive the cash flow constraint, assume that the individual receives current income $v(t)$ and a return on capital $ik(t)$ where $k(t)$ is the level of capital invested. Income is then used for consumption $c(t)$ and change in investment $k'(t)$. Thus, the cash flow equation becomes

$$v(t) + ik(t) = c(t) + k'(t) \quad (6.66)$$

with boundary conditions $k(0) = k_0$ and $k(T) = k_T$. The solution is derived by

$$c(t) = v(t) + ik(t) - k'(t) \Rightarrow \max \int_0^T e^{-rt} U(v(t) + ik(t) - k'(t)) dt \quad (6.67)$$

The Euler equation is then formed as

$$f(t, k(t), k'(t)) = e^{-rt} U(v(t) + ik(t) - k'(t)) \Rightarrow \begin{cases} f_k = ie^{-rt} U'(c(t)) \\ f_{k'} = -e^{-rt} U'(c(t)) \end{cases} \quad (6.68)$$

Using the integral approach

$$\begin{aligned} \int_t^{t+\Delta} ie^{-rs} U'(c(s)) ds &= \int_t^{t+\Delta} \frac{d[-e^{-rs} U'(c(s))]}{ds} ds \\ \int_t^{t+\Delta} ie^{-rs} U'(c(s)) ds &= -e^{t+\Delta} U'(c(t+\Delta)) + e^{-rt} U'(c(t)) \quad (6.69) \\ e^{-rt} U'(c(t)) &= i \int_t^{t+\Delta} U'(c(s)) ds + e^{-r(t+\Delta)} U'(c(t+\Delta)) \end{aligned}$$

Interpretation

$$e^{-rt} U'(c(t)) \quad (6.70)$$

is the marginal utility of consuming an additional dollar at time t . While

$$i \int_t^{t+\Delta} e^{-rs} U'(c(s)) ds + e^{-r(t+\Delta)} U'(c(t+\Delta)) \quad (6.71)$$

is the marginal value of the postponed dollar. Implicitly,

$$i \int_t^{t+\Delta} e^{-rs} ds \quad (6.72)$$

is the small change in dollars resulting from investing the dollar for consumption later rather than consuming it in the current period. Thus,

$$i \int_t^{t+\Delta} e^{-rs} U'(c(s)) ds \quad (6.73)$$

is the change in utility from investing the dollar of consumption and receiving the consumption value in the next period. Similarly,

$$U'(c(t + \Delta)) e^{-r(t+\Delta)} \quad (6.74)$$

is the marginal value of consumption at the next increment in time.

Back to the Euler equation

$$\begin{aligned} \frac{d}{dt} [-e^{-rt} U'(c(t))] &= r e^{-rt} U'(c(t)) - e^{-rt} U''(c(t)) c'(t) \\ \therefore i e^{-rt} U'(c(t)) &= r e^{-rt} U'(c(t)) - e^{-rt} U''(c(t)) c'(t) \\ (i - r) U'(c(t)) &= -U''(c(t)) c'(t) \\ i - r &= \frac{U''(c(t)) c'(t)}{U'(c(t))} \end{aligned} \quad (6.75)$$

Given that $-U''/U' > 0$, $C'(t) > 0$ if $i > r$. Remember that i is the return on capital and r is the impatience. Thus, you accumulate capital if the interest rate is greater than the impatience rate. Letting $U(c(t)) = \ln(c(t))$

$$\begin{aligned} U'(c(t)) &= \frac{1}{c(t)} \\ U''(c(t)) &= -\frac{1}{c(t)^2} \\ -\frac{U''(c(t)) c'(t)}{U'(c(t))} &= \frac{c'(t)}{c(t)} = i - r \Rightarrow c(t) = c(0) e^{(i-r)t} \end{aligned} \quad (6.76)$$

the concept of an excessive burden as the loss of consumer surplus due to the inefficiency imposed by the tax. [Levhari and Sheskinski (1972)] Levhari, D. Sheskinski, E. develop a new, dynamic idea of excess burden based on intertemporal maximization of utility. Thus, we define the excess burden of a tax as the loss of consumer's surplus summed, with the appropriate discounting, over the lifetime of the individual. In principle, the [Harberger (1964)] Harberger, A. C. formula may also be applied to the intertemporal case, interpreting some of the different goods as the same goods supplied in different time periods. There are, however, some special features of the intertemporal case which require elaboration. The intertemporal case may involve an infinite number of goods (when time is treated continuously or with an open horizon). The imposition of a tax on a particular good involves the quantities of the same good in all future time periods; that is, it is equivalent to imposition of a set of taxes on numerous goods at the same rate. There are difficulties involved in the calculation of the intertemporal "expenditure function" (required for the Harberger formula).

6.3.4.1 Derivation of the Lifetime Program

In the pure capital model, income is defined as

$$y(t) = r(1 - \lambda)a(t) + v(t) \quad (6.77)$$

where $y(t)$ is the income at time t composed of net profits and government transfers, $a(t)$ is the level of capital stock at time t , r is the interest rate, λ is the tax rate $0 < \lambda < 1$, and $v(t)$ is the level of government payments. Income is divided into two components: new investment and consumption

$$y(t) = c(t) + \dot{a}(t) \quad (6.78)$$

Assuming a utility function

$$U(c(t)) = \ln(c(t)) \quad (6.79)$$

the welfare function then becomes

$$V = \int_0^T e^{-\delta t} U(c(s)) ds \quad (6.80)$$

($\delta > 0$).

Given an initial stock of capital at time zero, the individual maximizes V subject to a terminal solvency constraint

$$a_T = 0 \quad (6.81)$$

Thus,

$$\begin{aligned} V &= \int_0^T \ln(r(1 - \lambda)a(s) + v(s) - a'(s)) e^{-\delta s} ds \\ f(t, a, a') &= \ln(r(1 - \lambda)a(t) + v(t) - a'(t)) e^{-\delta t} \\ \Rightarrow \begin{cases} f_a &= \frac{r(1 - \lambda)e^{-\delta t}}{(r(1 - \lambda)a(t) + v(t) - a'(t))} \\ f_{a'} &= \frac{-e^{-\delta t}}{(r(1 - \lambda)a(t) + v(t) - a'(t))} \end{cases} \end{aligned} \quad (6.82)$$

Differentiating the partial with respect to a' with respect to time

$$\begin{aligned} \frac{df_{a'}}{dt} &= \frac{\delta e^{-\delta t}}{(r(1 - \lambda)a(t) + v(t) - a'(t))} \\ &+ \frac{(r(1 - \lambda)a'(t) + v'(t) - a''(t))e^{-\delta t}}{(r(1 - \lambda)a(t) + v(t) - a'(t))^2} \end{aligned} \quad (6.83)$$

Substituting for consumption and the differential of consumption yields a Euler equation

$$\begin{aligned}\frac{r(1-\lambda)e^{-\delta t}}{c(t)} &= \frac{\delta e^{-\delta t}}{c(t)} + \frac{e^{-\delta t}c'(t)}{[c(t)]^2} \\ [r(1-\lambda) - \delta] - \frac{c'(t)}{c(t)} &= 0 \\ c(t) &= c_0 e^{[r(1-\lambda) - \delta]t}\end{aligned}\tag{6.84}$$

Assuming that $v(t) = \lambda r a(t)$, or that government returns all tax revenues to the individual, from Equations 6.77 and 6.78

$$\begin{aligned}y(t) &= r(1-\lambda)a(t) + \lambda r a(t) \\ &= r a(t) - \lambda r a(t) + \lambda r a(t) \\ &= r a(t) \\ \therefore a'(t) - r a(t) &= -c(t)\end{aligned}\tag{6.85}$$

As we previously discussed, this type of differential is solved by multiplying both sides by e^{-rs} and integrating both sides

$$a(t)e^{-rt} = a(0) - \int_0^t c(s)e^{-rs}ds\tag{6.86}$$

Using the terminal condition

$$\begin{aligned}a(T)e^{-rT} &= a(0) - \int_0^T c(s)e^{-rs}ds \\ a(T) = 0 &\Rightarrow a(0) = \int_0^T c(s)e^{-rs}ds \\ &= \int_0^T c(0)e^{[r(1-\lambda) - \delta]s - rs}ds \\ &= \int_0^T c(0)e^{-[r\lambda + \delta]s}ds \\ &= \frac{c(0)e^{-[r\lambda + \delta]T}}{-[r\lambda + \delta] - \frac{c(0)}{-[r\lambda + \delta]}} \\ &= c(0) \frac{[1 - e^{-[r\lambda + \delta]T}]}{r\lambda + \delta}\end{aligned}\tag{6.87}$$

Reversing the equality

$$c(0) = a(0) \frac{\lambda r + \delta}{1 - e^{-[\lambda r + \delta]T}} \quad (6.88)$$

thus

$$c(t, \lambda, a_0) = a_0 \frac{\lambda r + \delta}{1 - e^{-[\lambda r + \delta]T}} e^{[r(1-\lambda) - \delta]t} \quad (6.89)$$

6.3.4.2 Fish Example

One application of calculus of variations is the design of fishing policies. Typically, the change in the fish population is modeled as some function of current fish stock. For example, using the logistic function

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{k}\right) \quad (6.90)$$

where r is the intrinsic growth rate and k is the carrying capacity. Adding fishing to this equation of motion

$$\Rightarrow \begin{cases} \frac{dx}{dt} = rx \left(1 - \frac{x}{k}\right) - qEx \\ F(x) = rx \left(1 - \frac{x}{k}\right) \text{ natural reproduction} \\ h(t) = qEx \text{ harvest} \end{cases} \quad (6.91)$$

To integrate these into a model of fishing we construct a profit function for each unit of effort

$$R\Delta t = [ph(t) - cE(t)] \Delta t \quad (6.92)$$

where R is the revenue for the increment in time Δt , p is the price of the output, c is the cost of effort and $E(t)$ is the level of effort. Rewriting the harvest variable as a function of effort so that $h(t) = G(x(t))$ we get

$$\begin{aligned} R\Delta t &= [pG(x(t)) - c] E(t) \Delta t \\ &= [p - c(x(t))] h(t) \Delta t \\ c(x(t)) &= \frac{c}{G(x(t))} \\ G(x(t)) &= \frac{h(t)}{E(t)} \end{aligned} \quad (6.93)$$

Therefore, the present value of a harvest scenario becomes

$$PV = \int_0^{\infty} e^{-\delta t} [p - c(x(t))] h(t) dt \quad (6.94)$$

To integrate the equation of motion, we solve the fish stock equation backward yielding

$$\begin{aligned} \frac{dx}{dt} &= x'(t) = F(x(t)) - h(t) \\ h(t) &= F(x(t)) - x'(t) \end{aligned} \quad (6.95)$$

Therefore the planning problem becomes

$$\begin{aligned} PV &= \int_0^{\infty} e^{-\delta t} [p - c(x(t))] (F(x(t)) - x'(t)) dt \\ f(t, x, x') &= e^{-\delta t} [p - c(x(t))] (F(x(t)) - x'(t)) \\ \Rightarrow \begin{cases} f_x = e^{-\delta t} \{ c'(x(t)) (F(x(t)) - x'(t)) + \\ [p - c(x(t))] F'(x(t)) \} \\ f_{x'} = -e^{-\delta t} [p - c(x(t))] \end{cases} \\ e^{-\delta t} \{ -c'(x(t)) (F(x(t)) - x'(t)) + [p - c(x(t))] F'(x(t)) \} &= \\ e^{-\delta t} \{ \delta [p - c(x(t))] + c'(x(t)) x'(t) \} \end{aligned} \quad (6.96)$$

Solving the Euler equation

$$\begin{aligned} -c'(x(t)) F(x(t)) + c'(x(t)) + (p - c(x(t))) F'(x(t)) &= \\ \delta [p - c(x(t))] + c'(x(t)) x'(t) & \\ -\frac{c'(x(t)) F(x(t))}{p - c(x(t))} + F'(x(t)) &= \delta \end{aligned} \quad (6.97)$$

In some cases, the solution of the differential equation implied by the Euler equation can be simplified by the structure of the problem. Particularly, $f(t, x, x')$ if t , x , or x' are zero, then the differential equation implied by Euler is simplified.

Example 6.5. f depends only on t and x' ($f = f(t, x')$). Following the calculus of variations

$$f_x = 0 = \frac{df_{x'}}{dt} \quad (6.98)$$

If $df_{x'}/dt = 0$ then $f_{x'}$ is a constant. For example

$$\begin{aligned}
& \int_{t_0}^{t_1} \left(3x'(t) - t[x'(t)]^2 \right) dt \\
& f_{x'} = 3 - 2tx'(t) = c_0 \\
& -2tx'(t) = c_0 - 3 \\
& x'(t) = \frac{c_0 - 3}{2t} = \frac{c_1}{t}
\end{aligned} \tag{6.99}$$

Example 6.6. F depends on x and x'

$$\begin{aligned}
f_x &= f_{x'x}x' + f_{x'x'}x'' \\
f_x - f_{x'x}x' - f_{x'x'}x'' &= 0
\end{aligned} \tag{6.100}$$

Multiplying by x'

$$\begin{aligned}
& x'[f_x - f_{x'x}x' - f_{x'x'}x''] = 0 \\
& f_x x' - [x'f_{x'x}x' + x'f_{x'x'}x''] = 0 \Leftrightarrow \begin{cases} f_x x' \Leftarrow \frac{df}{dt} \\ x'f_{x'x}x' + x'f_{x'x'}x'' \Leftarrow \frac{d[x'f_x]}{dt} \end{cases}
\end{aligned} \tag{6.101}$$

Thus,

$$\frac{d[f - x'f_x]}{dt} = 0 \Rightarrow f - x'f_x = c_0 \tag{6.102}$$

Example 6.7. f depends only on x' . The Euler equation is $f_{x'x'}x'' = 0$. Thus, along the extremal case either: (1) $f_{x'x'} = 0$ (in Kamien and Schwartz $f_{x'x'}(x') = 0$ because $f_{x'x'}$ is a function of x'), or (2) $x''(t) = 0$ which is identical to $x' = \text{const}$ which yields an extremal of the form $x(t) = c_1t + c_2$. If f is linear in $x' - f_{x'x'} = 0$, the Euler equation is an identity and any x satisfies it. If f depends only on x' , but is not linear the graphics of the extremals are straight lines

$$\begin{aligned}
& \int_{t_0}^{t_1} (x'(t))^3 e^{x'(t)} dt \\
& \text{s.t. } x(t_0) = x_0, x(t_1) = x_1 \\
& f = (x'(t))^3 e^{x'(t)} \Rightarrow \begin{cases} f_x = 0 \\ f_{x'} = 3(x'(t))^2 e^{x'(t)} + (x'(t))^3 e^{x'(t)} \end{cases}
\end{aligned} \tag{6.103}$$

The solution in this case must be of the form $x(t) = c_1t + c_2$.

Example 6.8. f depends on t and x only. The Euler equation becomes $f_x(t, x) = 0$ which is the static problem.

Example 6.9. f is linear in x' of the form $f = A(t, x) + B(t, x)x'$. The Euler equation is

$$A_x + B_x x' = B_t + B_x x' \Leftrightarrow \begin{cases} f_x = A_x + B_x x' \\ f_{x'} B(t, x) \end{cases} \quad (6.104)$$

$$A_x = B_t$$

This is not a differential equation. If the solution $x(t)$ of this equation satisfies the boundary conditions, it may be the optimal solution. Alternatively, $A_x = B_t$ may be an identity $A_x(t, x) = B_t(t, x)$, satisfied by any function x . Then according to the integrability theorem of differential equations, there is a function $P(t, x)$ such that $P_t = A$, $P_x = B$, $P_{tx} = A_x = B_t$ and

$$f(t, x, x') = A + Bx' = P_t + P_x x' = \frac{dP}{dt} \quad (6.105)$$

Thus, the integrand is the total derivative of P and

$$\int_{t_0}^{t_1} (A + Bx') dt = \int_{t_0}^{t_1} \left(\frac{dP}{dt} \right) dt = P(t_1, x(t_1)) - P(t_0, x(t_0)) \quad (6.106)$$

The value of the integral depends only on the endpoints; the path joining them is irrelevant. Any feasible path is optimal. This is analogous to the problem of maximizing a constant function; any feasible point will yield the same value.

Two integrands that appear different can lead to the same Euler equation and thus have the same extremals. This happens if the integrands differ by an exact differential

$$Q(t, x, x') \equiv \frac{dP}{dt} = P_t(t, x) + P_x(t, x)x' \quad (6.107)$$

then for any twice differentiable function $f(t, x, x')$, the two integrals

$$\int_{t_0}^{t_1} f(t, x, x') dt \text{ and } \int_{t_0}^{t_1} [f(t, x, x') + Q(t, x, x')] dt \quad (6.108)$$

Only differ by a constant $P(t_1, x_1) - P(t_0, x_0)$ and the Euler equations are identical.

Example 6.10 (Discounted Profit). This example can be viewed as a simplification of [Jorgenson (1973)] Jorgenson, D. W.. The general goal is to design an investment plan, $I(t)$, that maximizes the discounted value of profit. This investment is subject to a production process where the level of production is based on the level of capital, $K(t)$. In Jorgenson's original article, the production function also included a labor input, $L(t)$. The calculus of variations problem is then written as

$$\begin{aligned} & \int_0^T e^{-rt} [p(t) f(K(t) - c(t) I(t))] dt \\ & \text{s.t. } K(0) = K_0, \quad K(T) = K_T \\ & \quad I(t) = K'(t) + bK(t) \end{aligned} \quad (6.109)$$

Substituting the investment equation into the objective function yields

$$\begin{aligned} & \int_0^T e^{-rt} [p(t) f(K(t) - c(t) (K'(t) + bK(t)))] dt \\ & F = e^{-rt} [p(t) f(K(t) - c(t) (K'(t) + bK(t)))] \\ & \Rightarrow \begin{cases} F_K = [p(t) f'(K(t) - c(t)b)] e^{-rt} \\ F_{K'} = -c(t) e^{-rt} \end{cases} \\ & \Rightarrow e^{-rt} [p(t) f'(K(t) - c(t)b)] = e^{-rt} [rc(t) - c'(t)] \end{aligned} \quad (6.110)$$

Note that this is not a differential equation in $K(t)$. The optimal capital choice is separable.

6.3.5 Second Order Conditions

From traditional calculus, we are accustomed to two sets of criteria for maximizing a function: First Order Conditions $f'(x^*) = 0$ and Second Order Conditions $f''(x^*) < 0$. A parallel exists for the calculus of variations. We previously concluded that given a small variation around an optimal path

$$g(a) = \int_{t_0}^{t_1} f(t, x(t) + ah(t), x'(t) + ah'(t)) dt \quad (6.111)$$

we must have

$$g'(0) = \int [f_x h + f_{x'} h'] dt = 0 \quad (6.112)$$

This expression is called the first variation. The analog of the second derivative is the second variation

$$g''(0) = \int_{t_0}^{t_1} \left[f_{xx} h^2 + 2f_{xx'} h h' + f_{x'x'} (h')^2 \right] dt \quad (6.113)$$

If $x^*(t)$ maximizes the integrand, it is necessary that the second variation be nonpositive for all admissible functions h . The second partials evaluated along $(t, x^*(t), x'^*(t))$. This second order condition leads to the Legendre condition. The second order condition will surely be met if f is concave in x and x'

$$\begin{bmatrix} \Delta_1 & \Delta_2 \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xx'} \\ f_{x'x} & f_{x'x'} \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} \leq 0 \quad (6.114)$$

Indeed, if the integrand $f(t, x, x')$ is jointly concave in its second and third arguments, and $x^*(t)$ satisfies the Euler equation $f_x = df_{x'}/dt$, then $x^*(t)$ maximizes the integrand.

Proof. Suppose $f = (t, x, x')$ and $f^* = (t, x^*, x'^*)$. Defining $h(t) = x(t) - x^*(t)$ and $h'(t) = x'(t) - x'^*(t)$. Since f is concave, it follows that

$$\int_{t_0}^{t_1} [f - f^*] dt \leq \int_{t_0}^{t_1} [(x - x^*) f_x^* + (x' - x'^*) f_{x'}^*] dt \quad (6.115)$$

Replacing $x - x^*$ with h and $x' - x'^*$ with h'

$$\begin{aligned} \int_{t_0}^{t_1} [f - f^*] dt &\leq \int_{t_0}^{t_1} [f_x^* h + f_{x'}^* h'] dt \\ &\leq \int_{t_0}^{t_1} f_x^* h dt + \int_{t_0}^{t_1} f_{x'}^* h' dt \end{aligned} \quad (6.116)$$

Integrating the second term on the right hand side by parts yields

$$\begin{aligned} \int_{t_0}^{t_1} f_{x'}^* h' dt &= f_{x'}^*(t_1, x(t_1), x'(t_1)) h(t_1) - f_{x'}^*(t_0, x(t_0), x'(t_0)) h(t_0) \\ &\quad - \int_{t_0}^{t_1} \frac{df_{x'}^*}{dt} h dt \end{aligned} \quad (6.117)$$

Substituting this expression back into the preceding result yields Euler's equation which indicates that if the function is jointly concave in x and x' , then the second order conditions for maximization are met.

Unfortunately, for many problems of interest the function will not be concave in x and x' . The Legendre condition requires that the integrand be locally concave in x' along the optimal path. Returning to

$$g''(0) = \int_{t_0}^{t_1} \left[f_{xx} h^2 + 2f_{xx'} h h' + f_{x'x'} (h')^2 \right] dt \quad (6.118)$$

Focusing on the $2f_{xx'} h h'$ term and integrating by parts

$$\left. \begin{aligned} u = f_{xx'} &\Rightarrow du = \frac{df_{xx'}}{dt} \\ dv = 2hh'dt &\Rightarrow v = h^2 \end{aligned} \right\} \Rightarrow \int_{t_0}^{t_1} 2f_{xx'} h h' dt = \quad (6.119)$$

$$f_{xx'} h^2 \Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{df_{xx'}}{dt} h^2 dt$$

The second variation then becomes

$$g''(0) = \int_{t_0}^{t_1} \left[\left(f_{xx} - \frac{df_{xx'}}{dt} \right) h^2 + f_{x'x'} (h')^2 \right] dt. \quad (6.120)$$

□

Lemma 6.1. *Let $P(t)$ and $Q(t)$ be given continuous functions on $[t_0, t_1]$ and let the quadratic functional*

$$\int_{t_0}^{t_1} \left[P(t) (h'(t))^2 + Q(t) (h(t))^2 \right] dt \quad (6.121)$$

be defined for all continuous differentiable functions $h(t)$ on $[t_0, t_1]$ such that $h(t_0) = h(t_1) = 0$. A necessary condition for the integrand to be nonpositive for all such h is that $p(t) \leq 0$, $t_0 \leq t \leq t_1$. This lemma implies that

$$g''(0) = \int_{t_0}^{t_1} \left[\left(f_{xx} - \frac{df_{xx'}}{dt} \right) h^2 + f_{x'x'} (h')^2 \right] dt \leq 0 \quad (6.122)$$

if $f_{x'x'}(t, x^*(t), x^{*'}(t)) \leq 0$

This is the Legendre condition: a maximizing $x^*(t)$ must not only satisfy the Euler equation, but also the Legendre condition. A minimizing path must satisfy $f_{x'x'}(t, x^*(t), x^{*'}(t)) \geq 0$.

6.3.6 Isoperimetric Problem

General statement of the isoperimetric problem

$$\begin{aligned} & \max \int_{t_0}^{t_1} f(t, x, x') dt \\ & \text{s.t.} \int_{t_0}^{t_1} g(t, x, x') dt \\ & x(t_0) = x_0, x(t_1) = x_1. \end{aligned} \quad (6.123)$$

for the specific case

$$\begin{aligned} & \max \int_0^T e^{-rt} P(x) dt \\ & \text{s.t.} \int_0^T x dt = B \end{aligned} \quad (6.124)$$

The basis of this problem is that of an extractable resource where $x(t)$ is the rate of extraction and there is B units of resource to extract

$$\int_0^T x dt = y(t) \Rightarrow y'(t) = x(t) \quad (6.125)$$

Thus, a change in variables yields

$$\begin{aligned} & \max \int_0^T e^{-rt} P(y'(t)) dt \\ & \text{s.t.} y(0) = 0, y(T) = B \end{aligned} \quad (6.126)$$

However, another way to do the same thing is to use a Lagrange multiplier

$$\begin{aligned} & \max \int_0^T e^{-rt} P(x) dt + \lambda \left[B - \int_0^T x dt \right] \\ & = \int_0^T [e^{-rt} P(x) - \lambda x] dt + \lambda B \\ & \quad f_x = e^{-rt} P'(x) - \lambda = 0 \\ & \quad e^{-rt} P'(x) = \lambda \end{aligned} \quad (6.127)$$

In a more general case

$$\begin{aligned} & \int_{t_0}^{t_1} [f(t, x, x') - \lambda g(t, x, x')] dt - \lambda B \\ & F_x = f_x - \lambda g_x = \frac{d[f_{x'} - \lambda g_{x'}]}{dt} \end{aligned} \quad (6.128)$$

Example 6.11. Isoperimetric formulation

$$\begin{aligned}
& \min \int_0^1 [x'(t)]^2 dt \\
& \text{s.t. } \int_0^1 x(t) dt = B \\
& x(0) = 0, x(1) = 2
\end{aligned} \tag{6.129}$$

Writing the augmented problem

$$\int_0^1 \left([x'(t)]^2 - \lambda x(t) \right) dt + \lambda B \tag{6.130}$$

The Euler equation is then

$$\begin{aligned}
2x''(t) + \lambda &= 0 \\
x''(t) &= -\frac{\lambda}{2} \\
x'(t) &= -\frac{\lambda t}{2} + c_1 \\
x(t) &= -\frac{\lambda t^2}{4} + c_1 t + c_2
\end{aligned} \tag{6.131}$$

We need to determine three constants $-\lambda$, c_1 , c_2 using the integral constraint and the boundary conditions. First using the endpoints

$$\begin{aligned}
x(0) &= -\frac{\lambda(0)^2}{4} + c_1(0) + c_2 = 0 \Rightarrow c_2 = 0 \\
x(1) &= -\frac{\lambda 1^2}{4} + c_1(1) = 2 \Rightarrow c_1 = 2 + \frac{\lambda}{4}
\end{aligned} \tag{6.132}$$

Next, using the integral condition

$$\begin{aligned}
\int_0^1 \left(-\frac{\lambda t^2}{4} + \left[2 + \frac{\lambda}{4} \right] t \right) dt &= -\frac{\lambda(1)^3}{12} + \left(1 + \frac{\lambda}{8} \right) (1) = B \\
\Rightarrow -\frac{\lambda}{12} + \frac{\lambda}{8} &= B - 1 \\
\Rightarrow \lambda &= 24(B - 1)
\end{aligned} \tag{6.133}$$

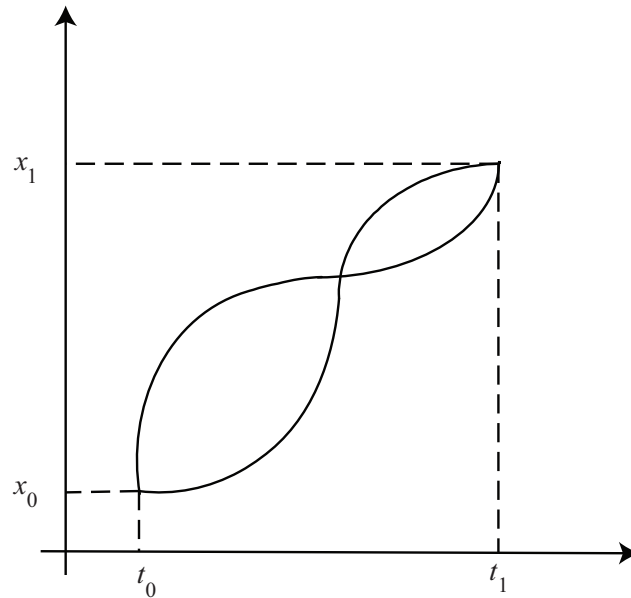


Fig. 6.5 Fixed Endpoint Problem

6.3.7 Free End Value Problems

To this point we have been interested in solving for the optimal path between two points $x(t_0)$ and $x(t_1)$ as depicted in Figure 6.5. Now the focus shifts to the case that we know the initial point, but do not specify the terminal point. We are simply searching for the path that crosses a certain point in time as depicted in Figure 6.6. Mathematically, the problem becomes

$$\begin{aligned} \max \int_{t_0}^{t_1} f(t, x(t), x'(t)) dt \\ \text{s.t. } x(t_0) = x_0 \end{aligned} \quad (6.134)$$

Back in the primordial past, we applied Liebnitz's rule to

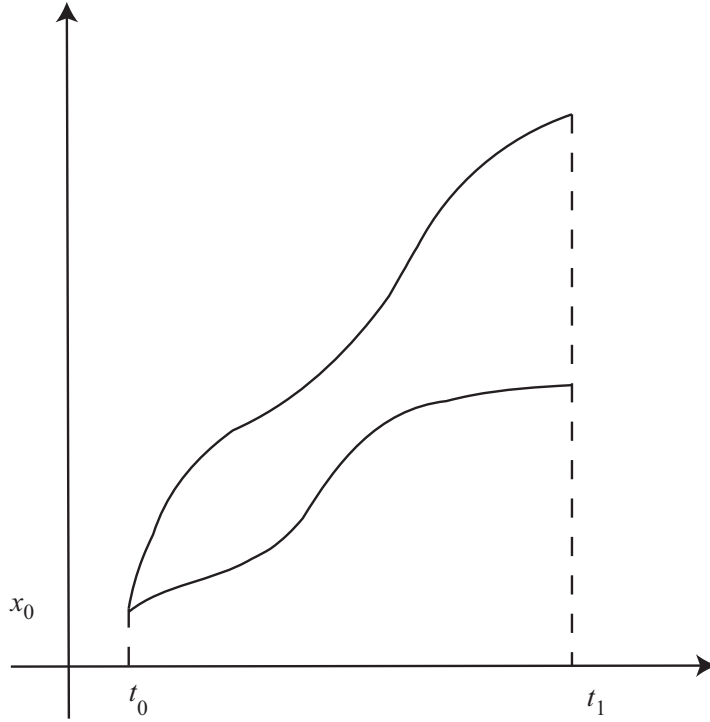


Fig. 6.6 Free Endpoint Problem

$$\begin{aligned}
 \frac{d}{d\epsilon} \int_{t_0}^{t_1} f(t, \Phi(\epsilon), \Phi'(\epsilon)) dt &= \frac{dt_1}{d\epsilon} f(t, \Phi(\epsilon), \Phi'(\epsilon)) - \frac{dt_0}{d\epsilon} f(t, \Phi(\epsilon), \Phi'(\epsilon)) \\
 &\quad + \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \Phi} \frac{d\Phi}{d\epsilon} + \frac{\partial f}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right] dt \\
 \frac{dt_1}{d\epsilon} &= \frac{dt_0}{d\epsilon} = 0
 \end{aligned} \tag{6.135}$$

Integrating the last term by parts then yielded

$$\begin{aligned}
 \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \Phi} \frac{d\Phi}{d\epsilon} - \frac{\partial f}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right] dt &= \\
 \left(\frac{\partial f}{\partial \Phi'} \phi(t) \right) \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \Phi} - \frac{d}{dt} \frac{\partial f}{\partial \Phi'} \right] \phi(t) dt
 \end{aligned} \tag{6.136}$$

Letting ϵ go to zero yields

$$f_{x'}(t_1, x(t_1), x'(t_1)) \phi(t_0) - f_{x'}(t_0, x(t_0), x'(t_0)) + \int_{t_0}^{t_1} \left[f_x - \frac{df_{x'}}{dt} \right] \phi(t) dt \quad (6.137)$$

In the current formulation, we will impose that $\phi(t_0) = 0$, but not $\phi(t_1)$. This yields the original Euler condition plus

$$f_{x'}(t_1, x(t_1), x'(t_1)) = 0 \quad (6.138)$$

6.3.8 Inequality Constraints

The basic setup is out standard dynamic problem

$$\begin{aligned} & \max \int_{t_0}^{t_1} F(t, x, x') dt \\ & \text{s.t. } x(t) \leq R(t), x(t_0) = x_0, x(t_1) = x_1 \end{aligned} \quad (6.139)$$

Thus, the constraint $x(t) \leq R(t)$ may or may not be binding at some point for the optimal path. Obviously, if the constraint is not binding, it does not affect the original formulation of the problem. If, on the other hand, the constraint is binding, it affects the optimal path as depicted in Figure 6.7. The optimal path is affected by the inequality constraint at t_1^* and follows the inequality constraint through point t_2^* at which point it diverges from the constraint. The question is then: Where does the optimal path join and leave the constraint? This problem can be formulated similarly to free endpoint problem.

Mathematically, the optimality conditions for this problem can be discussed within the context of a total variation. To define the total variation of the function, we return to the most basic derivation

$$\max \int_{a+\epsilon \delta a}^{b+\epsilon \delta b} F(t, \Phi, \Phi') dt \quad (6.140)$$

where δa denotes a variation in the lower bound of the integral (the starting time) and δb denotes a variation in the upper bound of the integral (the ending time). In this case, the scaling parameter for the variation ϵ is pulling double duty. It not only controls the perturbation of the optimal path in the Euler equation sense, but also controls the perturbation of either endpoint. By Taylor series expansion at the upper bound, we see that

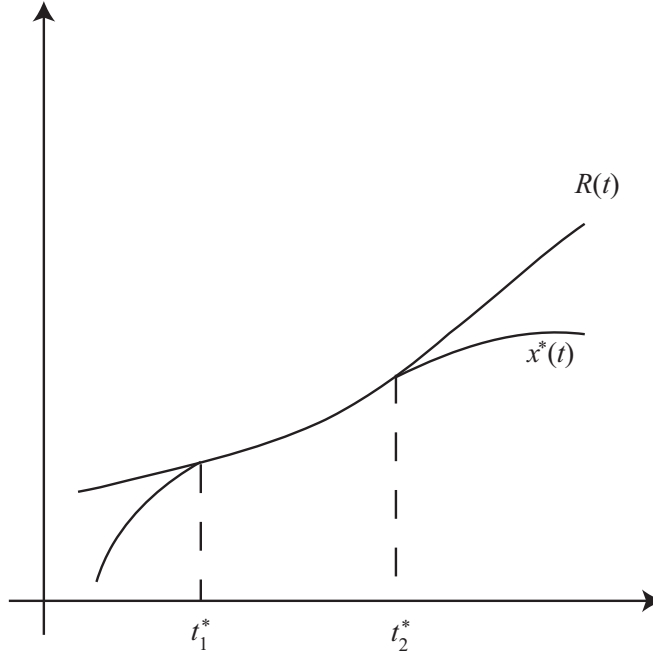


Fig. 6.7 Constrained Euler Solution

$$\begin{aligned} y(a + \epsilon \delta a) &\approx y(a) + y'(a)(a + \epsilon \delta a - a) \\ &= y(a) + y'(a)(\epsilon \delta a) \end{aligned} \quad (6.141)$$

Letting $y'(a)\epsilon$ go to $y^{*'}(a)$

$$\begin{aligned} y(a + \epsilon \delta a) &= y(a) + y^*(a) \delta a \\ \therefore \delta y_a &= y(a) + y^{*'}(a) - y^*(a) \\ \delta y_a &= \phi(x) + y^{*'}(a) \delta a \quad \exists: \phi(x) = y(a) - y^*(a) \\ \phi(x) &= \delta y_a - y^{*'}(a) \delta a \end{aligned} \quad (6.142)$$

The perturbation along the feasible path then becomes

$$\begin{aligned} \frac{dI}{d\epsilon} &= \delta b F - \delta a F + \frac{\partial F}{\partial \Phi'} (\delta y_b - y^{*'} \delta b) - \frac{\partial F}{\partial \Phi'} (\delta y_a - y^{*'} \delta a) + \\ &\quad \int_a^b \left(\frac{\partial F}{\partial \Phi} \frac{d\Phi}{d\epsilon} - \frac{\partial F}{\partial \Phi'} \frac{d\Phi'}{d\epsilon} \right) dt \end{aligned} \quad (6.143)$$

Letting ϵ go to zero and collecting terms

$$\delta I = (F - x'F_{x'})|_b \delta b + F_{x'}|_b \delta x(b) - (F - x'F_{x'})|_a \delta a - F_{x'}|_a + \int_a^b \left(F_x - \frac{dF_{x'}}{dt} \right) h dt \quad (6.144)$$

This problem is sufficient to discuss the problem of joining or leaving the optimal path. The question of joining the constraint is the case where $\delta b \neq 0$ and $\delta x(b) \neq 0$ and the leaving constraint $\delta a \neq 0$ and $\delta x(a) \neq 0$. Along the constraint, intuitively, $\delta I \leq 0$. Thus, either the Euler condition is less than or equal to zero, or the path is inadmissible. Solving the problems in sections, if the optimal path joins the constraint then

$$\int_{t_0}^{t_1} F(\cdot) dt \Rightarrow F_{x'} \delta x_1 + (F x' F_{x'}) \delta t_1 \quad (6.145)$$

at t_1 . At this condition $R(t_1) \delta t_1 = \delta x_1$, thus

$$[F + F_{x'}(R' - x')] \delta t_1. \quad (6.146)$$

This is compared with the value along the constraint

$$\int_{t_1+\delta t_1}^{t_1} F(t, R(t), R'(t)) dt \approx -F(t_1, R(t_1), R'(t_1)) \delta t_1. \quad (6.147)$$

Combining the two conditions

$$[F(t, x, x') + F_{x'}(t, x, x')(R' - x') - F(t, R, R')]|_{t_1} \delta t_1. \quad (6.148)$$

Yielding the condition that

$$F(t, x, x') + F_{x'}(t, x, x')(R' - x') - F(t, R, R') = 0 \text{ at } t_1 \quad (6.149)$$

6.4 Optimal Control

6.4.1 Basic Formulation

The basic formulation of the optimal control problem is

$$\begin{aligned} & \max \int_{t_0}^{t_1} f(t, x(t), u(t)) dt \\ & \text{s.t. } x'(t) = g(t, x(t), u(t)) \\ & t_0, t_1, x(t_0) = x_0 \text{ fixed, } x(t_1) \text{ free} \end{aligned} \quad (6.150)$$

where $x(t)$ is a *state variable* state variable and $u(t)$ is a *control variable* control variable. The typical calculus of variations problem can be reformulated as a control problem

$$\left\{ \begin{array}{l} \max \int_{t_0}^{t_1} f(t, x(t), x'(t)) dt \\ \text{s.t. } x(t_0) = x_0 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \max \int_{t_0}^{t_1} f(t, x(t), u(t)) dt \\ u(t) = x'(t) \\ x(t_0) = x_0 \end{array} \right\} \quad (6.151)$$

Rewriting Equation 6.150 in Lagrange form yields

$$\begin{aligned} J &= \int_{t_0}^{t_1} [f(t, x(t), u(t)) + \lambda(t)(g(t, x(t), u(t)) - x'(t))] dt \\ &= \int_{t_0}^{t_1} [f(t, x(t), u(t)) + \lambda(t)g(t, x(t), u(t)) - \lambda(t)x'(t)] dt \end{aligned} \quad (6.152)$$

Integrating the last term in the integrand by parts

$$\begin{aligned} \int_{t_0}^{t_1} \lambda(t)x'(t) dt &\Leftarrow \begin{cases} v = \lambda(t) \Rightarrow dv = \lambda'(t) dt \\ du = x'(t) dt \Rightarrow u = x(t) \end{cases} \\ \int_{t_0}^{t_1} \lambda(t)x'(t) dt &= \lambda(t_1)x(t_1) - \lambda(t_0)x(t_0) - \int_{t_0}^{t_1} \lambda'(t)x(t) dt \end{aligned} \quad (6.153)$$

Substituting the result of Equation 6.153 into the Lagrangian yields

$$\begin{aligned} J &= \int_{t_0}^{t_1} [f(t, x(t), u(t)) + \lambda(t)g(t, x(t), u(t)) + x(t)\lambda'(t)] dt \\ &\quad - \lambda(t_1)x(t_1) + \lambda(t_0)x(t_0) \end{aligned} \quad (6.154)$$

To develop the necessary conditions consider a variation around the optimal path $u(t)$

$$u(t) = u^*(t) + a h(t) \quad (6.155)$$

By construction, we can imagine a feasible path for $x(t)$ given the variation on the optimal path

$$x'(t) = g(t, x(t), u(t) + a h(t)) \quad (6.156)$$

This variation around $x(t)$ can then be written as a function of a

$$y(t, a) = \int x'(t) dt|_a = \int_{t_0}^t g(s, x(s), u(s) + ah(s)) ds \quad (6.157)$$

Just as in the calculus of variations, assume that there exists an optimal path for the state variable $x(t) \rightarrow x^*(t)$. If $a = 0$, $y(t, 0) = x^*(t)$. Further, the variation must be admissible so that $y(t_0, a) = x_0$ for any a .

Rewriting the objective function in terms of this alternative path yields

$$\begin{aligned} J(a) &= \int_{t_0}^{t_1} f(t, y(t, a), u^*(t) + ah(t)) dt \\ &= \int_{t_0}^{t_1} [f(t, y(t, a), u^*(t) + ah(t)) + \lambda(t) g(t, y(t, a), u^*(t) + ah(t)) \\ &\quad + y(t, a) \lambda'(t)] dt - \lambda(t_1) y(t_1, a) + \lambda(t_0) y(t_0, a) \end{aligned} \quad (6.158)$$

Differentiating with respect to a yields

$$J'(a) = \int_{t_0}^{t_1} [f_y y_a + f_u h + \lambda g_y y_a + \lambda' y_a] dt - \lambda(t_1) y_a \quad (6.159)$$

In this derivation $\lambda(t_0) y_a(t_0, a)$ must be zero because of the initial conditions on $y(t, a)$. Setting $a \rightarrow 0$ for the necessary conditions and collecting terms yields

$$J'(0) = \int_{t_0}^{t_1} [(f_x + \lambda g_x + \lambda') y_a + (f_u + \lambda g_u) h] dt - \lambda(t_1) y_a(t_1, 0) = 0 \quad (6.160)$$

For optimality

$$\begin{aligned} \int_{t_0}^{t_1} (f_x + \lambda g_x + \lambda') y_a dt = 0 &\Rightarrow f_x + \lambda g_x + \lambda' = 0 \\ &\Rightarrow f_x + \lambda g_x = -\lambda' \\ \int_{t_0}^{t_1} (f_u + \lambda g_u) h dt = 0 &\Rightarrow f_u + \lambda g_u = 0 \end{aligned} \quad (6.161)$$

and $\lambda(t_0) = 0$. In addition, λ must be chosen to satisfy the state equation

$$x'(t) = g(t, x(t), u(t)), \quad x(t_0) = x_0 \quad (6.162)$$

These conditions are the Pontryagin conditions with the Hamiltonian Hamiltonian

$$H(t, x(t), u(t), \lambda(t)) = f(t, x(t), u(t)) + \lambda(t) g(t, x(t), u(t)). \quad (6.163)$$

The Hamiltonian is then used to define: (1) The Optimality condition

$$\frac{\partial H}{\partial u} = f_u + \lambda g_u = 0 \Leftrightarrow \int_{t_0}^{t_1} (f_u + \lambda g_u) h dt = 0 \quad (6.164)$$

which depicts optimality with respect to the control variable. (2) the Multiplier condition

$$-\frac{\partial H}{\partial x} = -(f_x + \lambda g_x) = \lambda'(t) \Leftrightarrow \int_{t_0}^{t_1} (f_x + \lambda g_x + \lambda') y_a dt = 0 \quad (6.165)$$

which depicts the optimality with respect to the state variable. And (3) the Equation of Motion or state equation

$$\frac{\partial H}{\partial \lambda} = x'(t) = g(t, x(t), u(t)) \quad (6.166)$$

from differentiating $J(a)$ with respect to λ . In addition, $x(t_0) = x_0$ and $\lambda(t_1) = 0$ (the transversality condition).

In addition, the necessary conditions include

$$\begin{aligned} H_{uu}(t, x^*, u^*) &\leq 0 \text{ for a maximum, and} \\ H_{uu}(t, x^*, u^*) &\geq 0 \text{ for a minimum} \end{aligned} \quad (6.167)$$

Example 6.12. Show that optimal control yields the same solution as the calculus of variations. First, we transform the traditional calculus of variations into an optimal control problem as

$$\begin{aligned} \max \int_{t_0}^{t_1} f(t, x(t), u(t)) dt \\ \text{s.t. } x'(t) = u(t) \\ x(t_0) = x_0 \end{aligned} \quad (6.168)$$

First deriving the Pontryagin conditions

$$\begin{aligned} H &= f + \lambda u \\ \frac{\partial H}{\partial u} &= f_u + \lambda = 0 \\ -\frac{\partial H}{\partial x} &= -f_x = \lambda' \end{aligned} \quad (6.169)$$

Differentiating the optimality condition with respect to time yields

$$\frac{df_u}{dt} = -\lambda' \quad (6.170)$$

Substituting this result into the multiplier or costate equation yields

$$f_x = \frac{df_u}{dt} \quad (6.171)$$

Substituting in the equation of motion for the control variable then yields the Euler equation

$$f_x = \frac{df_{x'}}{dt} \quad (6.172)$$

To complete the demonstration, the costate condition at t_1 also implies that $f_{x'} = 0$ since $\lambda(t_1) = 0$.

Example 6.13. Consider the problem

$$\begin{aligned} \max \int_0^1 (x + u) dt \\ \text{s.t. } x'(t) = 1 - u^2, x(0) = 1 \end{aligned} \quad (6.173)$$

Forming the Hamiltonian

$$H = x + u + \lambda(1 - u^2) \Rightarrow \begin{cases} \frac{\partial H}{\partial u} = 1 + 2\lambda u = 0 \\ -\frac{\partial H}{\partial x} = -1 = \lambda' \\ \frac{\partial H}{\partial \lambda} = 1 - u^2 \end{cases} \quad (6.174)$$

Solving the Costate condition

$$\lambda(t) = 1 - t \quad (6.175)$$

The Optimality condition then implies that

$$\begin{aligned} 1 - 2(1 - t)u(t) &= 0 \\ u(t) &= \frac{1}{2(1 - t)} \end{aligned} \quad (6.176)$$

Finally, using the equation of motion

$$\begin{aligned} x'(t) &= 1 - \frac{1}{2(1 - t)^2} \\ x(t) &= t - \frac{1}{4(1 - t)} + \frac{5}{4} \end{aligned} \quad (6.177)$$

6.4.2 Sufficiency of Optimal Path

In the calculus of variations, the first and second order necessary conditions were also sufficient. That is if any solution obeyed the Euler equation and the Legendre condition, the path was optimum. The Legendre condition basically imposed concavity on x and x' . Similar conditions hold for the optimal control formulation. Assume that $x^*(t)$, $u^*(t)$, and $\lambda(t)$ for the optimum control formulation from Equation 6.150 satisfy the optimality conditions in Equations 6.164, 6.165, and 6.166 along with the transversality conditions

$$\lambda(t_1) = 0, \lambda(t) \geq 0. \quad (6.178)$$

These conditions imply that the optimal path is feasible. To establish sufficiency, we must establish

$$D \equiv \int_{t_0}^{t_1} (f^* - f) dt \geq 0 \quad (6.179)$$

If f is concave in $(x(t), u(t))$ then

$$D \geq \int_{t_0}^{t_1} [(x^*(t) - x(t)) f_x^* + (u^*(t) - u(t)) f_u^*] dt. \quad (6.180)$$

By the multiplier equation

$$f_x^* = -\lambda' - \lambda g_x \quad (6.181)$$

by the optimality condition

$$\begin{aligned} f_u^* &= \lambda g_u^* \\ \therefore D &\geq \int_{t_0}^{t_1} [(x^* - x)(-\lambda' - \lambda g_x) + (u^* - u)\lambda g_u^*] dt \end{aligned} \quad (6.182)$$

Focusing on $(x^* - x)(-\lambda' - \lambda g_x)$

$$\begin{aligned} &(-x^* \lambda' - x \lambda') + (x^* - x)(-\lambda' g_x) \\ \int_{t_0}^{t_1} x^* \lambda' dt &= x^* \lambda|_{t_0}^{t_1} - \int_{t_0}^{t_1} x^{*'} \lambda dt \\ &= x^*(t_0) \lambda(t_0) - \int_{t_0}^{t_1} g^* \lambda dt \end{aligned} \quad (6.183)$$

Substituting this result back into the expression for D yields

$$\begin{aligned} D &\geq \int_{t_0}^{t_1} [\lambda g^* - \lambda g - (x^* - x)(\lambda g_x) + (u^* - u)g_u] dt \\ &= \int_{t_0}^{t_1} \lambda [g^* - g - (x^* - x)g_x - (u^* - u)g_u] dt \end{aligned} \quad (6.184)$$

Since $\lambda \geq 0$

$$\begin{aligned} g^* - g - (x^* - x)g_x - (u^* - u)g_u &\geq 0 \\ \Rightarrow g^* - g &\geq (x^* - x)g_x + (u^* - u)g_u \end{aligned} \quad (6.185)$$

With the later condition being met if g is concave in x and u .

6.4.3 Interpretations

Consider

$$V(x_0, t_0) = \max \int_{t_0}^{t_1} f(t, x(t), u(t)) dt \text{ s.t. } x'(t) = g(t, x(t), u(t)) \quad (6.186)$$

remember

$$J = \int_{t_0}^{t_1} [f(t, x(t), u(t)) + \lambda(t)(g(t, x(t), u(t)) - x'(t))] dt \quad (6.187)$$

which is identical to the formulation that yielded the Pontryagin conditions

$$\therefore V(x_0, t_0) = \int_{t_0}^{t_1} [f(t, x^*(t), u^*(t)) + \lambda(t)g(t, x^*(t), u^*(t)) - \lambda(t)x^{*'}(t)] dt. \quad (6.188)$$

What we want to consider is a perturbation to the initial state $x_0 + a$ as depicted in Figure 6.8. Integrating by parts yields

$$V(x_0, t_0) = \int_{t_0}^{t_1} [f(t, x^*(t), u^*(t)) + \lambda(t)g(t, x^*(t), u^*(t)) - x^*(t)\lambda'(t)] dt - \lambda(t_1)x(t_1) + \lambda(t_0)x_0 \quad (6.189)$$

Evaluating this expression at $x_0 + a$ yields

$$\begin{aligned} V(x_0 + a, t_0) &= \int_{t_0}^{t_1} [f(t, x(t), u(t)) + \lambda(t)g(t, x(t), u(t)) - \lambda'(t)x(t)] dt \\ &\quad + \lambda(t_1)x(t_1) + \lambda(t_0)[x_0 + a] \end{aligned} \quad (6.190)$$

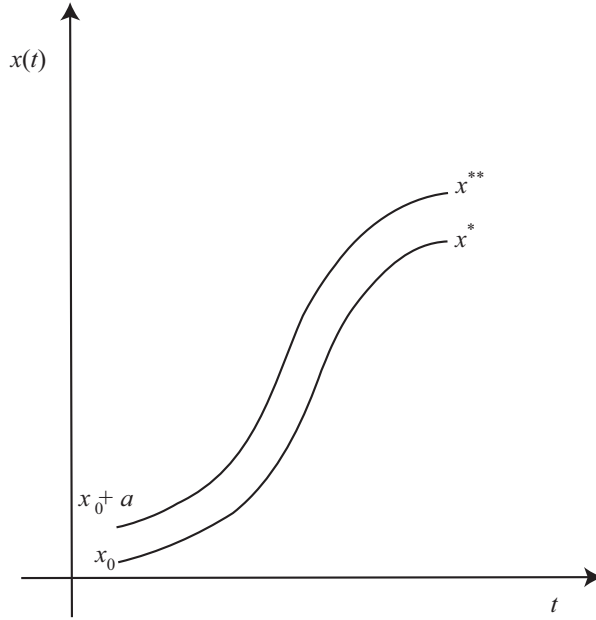


Fig. 6.8 Constant Perturbation in Optimal Path

The difference in the value functions can then be derived as

$$V(x_0 + a, t_0) - V(x_0, t_0) = \int_{t_0}^{t_1} [(f + \lambda g + \lambda' x) - (f^* + \lambda g^* + \lambda' x^*)] dt + \lambda(t_0)a - \lambda(t_1)[x(t_1) - x^*(t_1)] \quad (6.191)$$

Using a Taylor series expansion

$$(f + \lambda g + \lambda' x) - (f^* + \lambda g^* + \lambda' x^*) = (f_x^* + \lambda g_x^* + \lambda')(x - x^*) + (f_u^* + \lambda g_u^*)(u - u^*) + R_2 \quad (6.192)$$

Therefore

$$V(x_0 + a, t_0) - V(x_0, t_0) = \int_{t_0}^{t_1} [(f_x^* + \lambda g_x^* + \lambda')(x - x^*) + (f_u^* + \lambda g_u^*)] dt + \lambda(t_0)a - \lambda(t_1)[x(t_1) - x^*(t_1)] + \text{h.o.t} \quad (6.193)$$

Applying the optimality and multiplier conditions, and using the transversality condition that $\lambda(t_1) = 0$, we have

$$V(x_0 + a, t_0) - V(x_0, t_0) = \lambda(t_0) a + \text{h.o.t.} \quad (6.194)$$

By definition of the derivative

$$\frac{dV(x_0, t_0)}{da} = \lim_{a \rightarrow 0} \frac{\lambda(t_0) a}{a} = \lambda(t_0) \quad (6.195)$$

6.4.4 Several Variables

There is little change in the optimality conditions for the control problem of several variables. The only real change in complexity is the solution of a system of differential equations. Assume that we have a problem with two state variables and two control variables

$$\begin{aligned} \max \int_{t_0}^{t_1} f(t, x_1(t), x_2(t), u_1(t), u_2(t)) dt \\ \text{s.t. } x_1'(t) = g(t, x_1(t), x_2(t), u_1(t), u_2(t)) \\ x_2'(t) = g(t, x_1(t), x_2(t), u_1(t), u_2(t)) \\ x_1(t_0), x_2(t_0) \text{ fixed} \\ x_1(t_1), x_2(t_1) \text{ free} \end{aligned} \quad (6.196)$$

Again assume that we have optimal paths x_1^* , x_2^* , u_1^* , u_2^* around which we are going to form a variation. Assume

$$\begin{aligned} u_1(t) &= u_1^*(t) + ah_1(t) \\ u_2(t) &= u_2^*(t) + ah_2(t) \end{aligned} \quad (6.197)$$

Thus, we have a variation around the optimal path controlled by a . The deviation is independent through $h_i(t)$. Hence, there are two state paths consistent with the equations of motion

$$\begin{aligned} y_1(t, a) &= \int_{t_0}^{t_1} g_1(s, x_1(s), x_2(s), u_1(s), u_2(s)) ds \\ y_2(t, a) &= \int_{t_0}^{t_1} g_2(s, x_1(s), x_2(s), u_1(s), u_2(s)) ds \end{aligned} \quad (6.198)$$

such that $y_1(t_0, a) = x_1(t_0)$ and $y_2(t_0, a) = x_2(t_0)$.

Appendix A

Review of Linear Algebra

A.1 Canonical Form and Null Spaces

To demonstrate the meaning of null space matrices, I first want to discuss the use of canonical forms. A canonical form is a solution to an underidentified system of equations. For example in a crop selection model we may have a land and a labor constraint

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 100 \\ 2x_1 + 3x_2 + x_3 + 4x_4 + x_5 &= 250 \end{aligned} \quad (\text{A.1})$$

The matrix operations used to reduce this matrix to a row reduced form are

$$\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 100 \\ 2 & 3 & 1 & 4 & 1 & 250 \end{array} \right) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 100 \\ 0 & 1 & -1 & 2 & -1 & 50 \end{pmatrix} \quad (\text{A.2})$$

The operation in Equation A.2 subtracts 2 times row one from row two. The leading element of the second row of Equation A.2 is a 1 (or the next matrix operation would be to multiply by an identity matrix replacing the second element of the second row with the inverse of the leading term. However, given the result on the right hand side of Equation A.2 the goal is to subtract the second row from the first

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 100 \\ 0 & 1 & -1 & 2 & -1 & 50 \end{array} \right) = \begin{pmatrix} 1 & 0 & 2 & -1 & 2 & 50 \\ 0 & 1 & -1 & 2 & -1 & 50 \end{pmatrix} \quad (\text{A.3})$$

The result on the right hand side of Equation A.3 is a canonical form representing a solution to the system of equations. Specifically, $x_1 = 50$, $x_2 = 50$, $x_3 = 0$, $x_4 = 0$, and $x_5 = 0$ solves the system of equations. In

addition, the solution says something about maintaining feasibility in the nonbasic variables. Taking the columns from the nonbasic, or zero-level, variables we can form a homogeneous set of equations ($Ax = 0$)

$$\begin{aligned} 2x_3 - x_4 + 2x_5 &= 0 \\ -x_3 + 2x_4 - x_5 &= 0 \end{aligned} \quad (\text{A.4})$$

This system has the reduce form

$$\begin{aligned} x_3 + x_5 &= 0 \\ x_4 &= 0 \end{aligned} \quad (\text{A.5})$$

so any solution where $x_3 = -x_5$ maintains feasibility, or satisfies the equations.

A technical definition of the *null space* is the set of solutions of the linear equations $Ax = 0$. However, a fuller understanding of the null space involves a discussion of dimension. To define the system of linear equations more rigorously

$$Ax = b \Rightarrow \begin{cases} A \in M_{n \times m} \\ x \in M_{m \times 1} \\ b \in M_{n \times 1} \end{cases} \quad (\text{A.6})$$

where M denotes the set of matrices (or linear spaces). In the common parlance we have a system of n equations with m unknown values. In the most interesting null space $n < m$ or the system is underidentified (or not all the levels of the variables are determined by a given level of right hand sides $[b]$). The system is overidentified if $n > m$ which implies that the system is potentially inconsistent (i.e., not all the equations or constraints can be solved with the same set of variables $[x]$).

In order to "solve" this system (assuming that the rows of A are linearly independent) we know that we must have at least m nonzero variables in the canonical form. *The null space of the matrix A is the same dimension as the number of free variables.* Taking the first three variables from the above example

$$\begin{aligned} x_1 + x_2 + x_3 &= 100 \\ 2x_1 + 3x_2 + x_3 &= 250 \end{aligned} \quad (\text{A.7})$$

the reduced form for Equation A.7 is

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 50 \\ 0 & 1 & -1 & 50 \end{array} \right) \quad (\text{A.8})$$

Notice that any change

$$\begin{array}{rcl} \Delta x_1 & +2\Delta x_3 & = 0 \\ \Delta x_2 & -\Delta x_3 & = 0 \end{array} \quad (\text{A.9})$$

is still feasible. For example, if $\Delta x_3 = 10 \Rightarrow \Delta x_1 = -20$ and $\Delta x_2 = 10$. Then $x_3 + \Delta x_3 = 10$, $x_1 + \Delta x_1 = 30$, and $x_2 + \Delta x_2 = 60$ yieldin

$$\begin{array}{rcl} x_1 + x_2 + x_3 & = & 30 + 60 + 10 = 100 \\ 2x_1 + 3x_2 + x_3 & = & 60 + 180 + 10 = 250 \end{array} \quad (\text{A.10})$$

The null space is basically a line in 3 space which shows how the variables can be traded to maintain feasibility.

Expanding to the first four variables of the original model

$$\begin{array}{rcl} x_1 + x_2 + x_3 + x_4 & = & 100 \\ 2x_1 + 3x_2 + x_3 + 4x_4 & = & 250 \end{array} \quad (\text{A.11})$$

which yields a reduced form

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & -1 & 50 \\ 0 & 1 & -1 & 2 & 50 \end{array} \right) \quad (\text{A.12})$$

Again we note that

$$\begin{array}{rcl} \Delta x_1 & +2\Delta x_3 & -\Delta x_4 = 0 \\ \Delta x_2 & -\Delta x_3 & +2\Delta x_4 = 0 \end{array} \quad (\text{A.13})$$

In this case, the null space is a plane in 4 space. If $\Delta x_3 = 1$ and $\Delta x_4 = 5$ the

$$\begin{array}{rcl} \Delta x_1 & = & -2\Delta x_3 + \Delta x_4 = 3 \\ \Delta x_2 & = & \Delta x_3 - 2\Delta x_4 = -9 \end{array} \quad (\text{A.14})$$

which maintains feasibility.

A.2 Determinant

Just to make sure that you haven't picked up any bad habits, the determinant of any $n \times n$ matrix can be derived by expanding down any column or across any row of the matrix

$$\begin{aligned}
 \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} &= (-1)^{1+1} a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} + \\
 (-1)^{2+1} a_{21} \begin{vmatrix} a_{12} & a_{13} & a_{14} \\ a_{23} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} &+ (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} + \\
 (-1)^{4+1} \begin{vmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{23} \\ a_{32} & a_{33} & a_{34} \end{vmatrix} &
 \end{aligned} \tag{A.15}$$

where

$$\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{22} \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} - a_{32} \begin{vmatrix} a_{23} & a_{24} \\ a_{42} & a_{43} \end{vmatrix} + a_{42} \begin{vmatrix} a_{23} & a_{24} \\ a_{33} & a_{34} \end{vmatrix} \tag{A.16}$$

The eigenvalues of a matrix A are then defined as the solutions to the equation

$$|A - \lambda I| = 0 \tag{A.17}$$

where I is the identity matrix. One example of the use of eigenvalues is from differential equations

$$\begin{aligned}
 \frac{dv}{dt} &= 4v - 5w, \text{ where } v = 8 \text{ at } t = 0 \\
 \frac{dw}{dt} &= 2v - 3w, \text{ where } w = 8 \text{ at } t = 0
 \end{aligned} \tag{A.18}$$

Another way to write this problem is

$$u(t) = \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}, u_0 = \begin{pmatrix} 8 \\ 5 \end{pmatrix}, A = \begin{pmatrix} 4 & -5 \\ 2 & -2 \end{pmatrix}. \tag{A.19}$$

Thus,

$$\frac{du}{dt} = Au(t), u(0) = u_0. \quad (\text{A.20})$$

We are looking for solutions of the form

$$\begin{aligned} u(t) &= e^{\lambda t} y \\ w(t) &= e^{\lambda t} z. \end{aligned} \quad (\text{A.21})$$

Substituting this solution into the system of differential equations, we have

$$\begin{pmatrix} \lambda e^{\lambda t} y \\ \lambda e^{\lambda t} z \end{pmatrix} = \begin{pmatrix} 4e^{\lambda t} y - 5e^{\lambda t} z \\ 2e^{\lambda t} y - 3e^{\lambda t} z \end{pmatrix}. \quad (\text{A.22})$$

Next, we divide through by $e^{\lambda t}$ yielding

$$\begin{pmatrix} \lambda y \\ \lambda z \end{pmatrix} = \begin{pmatrix} 4y - 5z \\ 2y - 3z \end{pmatrix} \text{ or } \lambda \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \Rightarrow \lambda x = Ax \quad (\text{A.23})$$

which is the definition of the eigenvalues (λ) and the eigenvectors (x)

$$\begin{aligned} Ax - \lambda x &= 0 \\ (A - \lambda I)x &= 0 \end{aligned} \quad (\text{A.24})$$

where the eigenvalues of x lie in the null space of the matrix

$$A - \lambda I = \begin{pmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{pmatrix}. \quad (\text{A.25})$$

More intuitively, we know that the matrix

$$\begin{pmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{pmatrix} \quad (\text{A.26})$$

is singular by definition of the eigenvalue.

Given the existence of eigenvalues and eigenvectors, it is possible to derive the diagonal form of the matrix

$$S^{-1}AS = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \quad (\text{A.27})$$

where S is a matrix whose columns are the eigenvectors

$$\begin{aligned} S &= (x_1 \cdots x_n) \\ AS &= (\lambda_1 x_1 \cdots \lambda_n x_n) \\ AS &= (x_1 \cdots x_n) \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \end{aligned} \quad (\text{A.28})$$

Therefore, $S^{-1}S = I$, $AS = S\Lambda$, $S^{-1}AS = \Lambda$, and $A = S\Lambda S^{-1}$.

Notationally, in [Gill *et al.* (1981)] the eigenvalues of A are written as $\{\lambda_i[A]\}$. $P_i(A)$ is the product of the eigenvalues, and $\Pi(AB) = \Pi(A)\Pi(B)$. If A is symmetric (i) all the eigenvalues of A are real numbers, and (ii) the matrix has n distinct eigenvalues. If A is nonsingular then all its eigenvalues are nonzero and the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A . The maximum and minimum eigenvalues satisfy

$$\begin{aligned} \lambda_{\max}[A] &= \max_{x \neq 0} \frac{x'Ax}{x'x} \\ \lambda_{\min}[A] &= \min_{x \neq 0} \frac{x'Ax}{x'x} \end{aligned} \quad (\text{A.29})$$

A.3 Norms

A vector norm which will be denoted $\|x\|$, must satisfy three properties: (1) $\|x\| \geq 0$ for all x and $\|x\| = 0$ if and only if x is the zero vector, (2) for any real number δ , $\|\delta x\| = |\delta| \|x\|$, and (3) for any two vectors x and y , $\|x + y\| \leq \|x\| + \|y\|$. Our basic definition of the p -norm is

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}. \quad (\text{A.30})$$

From our discussion of the optimality conditions (Equation 2.11), we derived

$$f(x) - f(x^*) = \nabla_x f(x) dx + \frac{1}{2} dx \nabla_{xx}^2 f(x) dx' \leq 0 \quad (\text{A.31})$$

where $\nabla_{xx}^2 f(x)$ denotes the matrix of second derivatives or the hessian matrix. Equation A.31 implies for optimality

$$\|\nabla_x f(x)\|_p \rightarrow 0. \quad (\text{A.32})$$

The matrix norm is analogously defined with the additional property that (4) $\|AB\| \leq \|A\| \|B\|$.

Another value of the norm is a discussion of the stability of the matrix. For example, what are the properties of the solution of $Ax = b$ if we perturb the x vector?

$$\begin{aligned} A(x + \delta x) &= (b + \delta b) \\ x + \delta x &= A^{-1}b + A^{-1}\delta b. \end{aligned} \quad (\text{A.33})$$

Noting that $x = A^{-1}b$ we subtract x and take the norm of each side

$$\|\delta x\| = \|A^{-1}\| \|\delta b\|. \quad (\text{A.34})$$

Given that $\|b\| \leq \|A\| \|x\|$ by the original equation, we have

$$\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|} \quad (\text{A.35})$$

Multiplying both sides by the inequality involving the perturbation (δ), we get

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|}. \quad (\text{A.36})$$

This number $\|A\| \|A^{-1}\|$ is the condition number of A denoted $\text{cond}(A)$. Since

$$1 = \|I\| \leq \|A\| \|A^{-1}\| \quad (\text{A.37})$$

$\text{cond}(A) \geq 1$. If the condition number of A is large, the matrix is ill-conditioned. If the condition number is small, the matrix is well conditioned. Well conditioned systems are stable (i.e., small perturbations do not greatly affect the solution), but the solutions of ill-conditioned matrices tend to have large perturbations.

A.4 Matrix Factorizations

The matrix factorization $A = LU$ involves reducing a matrix into a lower diagonal matrix L and an upper diagonal matrix U . The LU factorization is computed by a set of elementary row operations. Taking the matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 6 & 2 & 6 \\ -4 & -3 & 9 \end{pmatrix}. \quad (\text{A.38})$$

The first series of row operations to reduce the matrix in Equation A.38 to an identity matrix involves subtracting 3 times row one from row two and adding 2 times row 1 to row 3. In matrix terms, this operation can be depicted as an the elemental matrix E_1 defined as

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad (\text{A.39})$$

The resulting matrix $A_2 = E_1 A_1$ becomes

$$A_1 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & -1 & 6 \\ 0 & -1 & 9 \end{pmatrix}. \quad (\text{A.40})$$

The second elemental matrix is then

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \quad (\text{A.41})$$

which yields the upper triangular matrix ($A_2 = E_2 A_1$)

$$A_2 = U = \begin{pmatrix} 2 & 1 & 0 \\ 0 & -1 & 6 \\ 0 & 0 & 3 \end{pmatrix}. \quad (\text{A.42})$$

The lower triangular matrix is then defined as

$$L = E_1^{-1} E_2^{-2} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}. \quad (\text{A.43})$$

The usefulness of the decomposition is twofold. First, the LU factorization is a quick way to solve large systems of equations. First, note that the solution of the system of linear equations $Ax = b$ for $b = (1 \ 5 \ 3)'$ is $x = (-5/2 \ 4 \ 1)'$. If we start with the augmented system of equations

$$(A|b) = \left(\begin{array}{ccc|c} 2 & 1 & 0 & 1 \\ 6 & 2 & 6 & 5 \\ -4 & -3 & 9 & 3 \end{array} \right) \quad (\text{A.44})$$

operating using the same sequence of elementary matrices used to generate the LU decomposition above

$$(A|b) = \left(\begin{array}{ccc|c} 2 & 1 & 0 & 1 \\ 0 & -1 & 6 & 2 \\ 0 & 0 & 3 & 3 \end{array} \right) \quad (\text{A.45})$$

In this system, the solution of the original equations is accomplished by taking the last equation first

$$\begin{aligned} 3x_3 &= 3 \\ x_3 &= 1 \end{aligned} \quad (\text{A.46})$$

Given that $x_3 = 1$ the next equation up in Equation A.45 becomes

$$\begin{aligned} -x_2 + 6(1) &= 2 \\ x_2 &= 4 \end{aligned} \quad (\text{A.47})$$

The first equation then completes the solution

$$\begin{aligned} 2x_1 + (4) &= 1 \\ x_1 &= -\frac{5}{2} \end{aligned} \quad (\text{A.48})$$

The second major use is in the cholesky decompositiy. If a matrix is symmetric and positive definite, it can be decomposed so that $A = PP'$. Using the matrix

$$A = \begin{pmatrix} 7 & 3 & 1 \\ 3 & 4 & 0 \\ 1 & 0 & 2 \end{pmatrix} \quad (\text{A.49})$$

First, we need to determine whether the matrix is positive definite. The eigenvalues of the matrix are 2.5572, 1.4841, and 8.9587 so the matrix is positive definite. Following the procedure above, we derive the L matrix as

$$L = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.4286 & 1.00000 & 0.00000 \\ 0.1429 & -0.1579 & 1.0000 \end{pmatrix} \quad (\text{A.50})$$

with the upper triangular matrix

$$U = \begin{pmatrix} 7.0000 & 3.0000 & 1.0000 \\ 0.0000 & 2.7143 & -0.4286 \\ 0.0000 & 0.0000 & 1.7895 \end{pmatrix} \quad (\text{A.51})$$

Noting that the diagonal of the L matrix is 1 and the diagonal of the U matrix are greater than zero, suppose that we constructed the matrix

$$D^{-1} = \begin{pmatrix} 0.1429 & 0.0000 & 0.0000 \\ 0.0000 & 0.3684 & 0.0000 \\ 0.0000 & 0.0000 & 0.5588 \end{pmatrix} \quad (\text{A.52})$$

which is one over the diagonal elements of the U matrix. Multiplying $D^{-1}U$ yields L' .

Appendix B

MINOS 5.6 Application

Text states that we need to develop Quadratic1.f and Quadratic1.SPECS)

Bibliography

- Featherstone, A. M. and Moss, C. B. (1994). Measuring economies of scale and scope in agricultural banking, *American Journal of Agricultural Economics* **76**, 3, pp. 655–661.
- Freund, R. F. (1953). The introduction of risk into a programming model, *Econometrics* **24**, 3, pp. 253–263.
- Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical Optimization* (Academic Press, New York).
- Harberger, A. C. (1964). The measurement of waste, *American Economic Review* **49**, pp. 134–146, <http://www.jstor.org/stable/1818490>.
- Irwin, G. and Baker, C. (1962). Effects of lender decisions on farm financial planning, Agricultural Experiment Station Bulletin 688, University of Illinois.
- Johnson, S., Tiefertiller, K., and Moore, D. (1967). Stochastic linear programming and feasibility problems in farm growth analysis, *Journal of Farm Economics* **49**, pp. 908–919.
- Jorgenson, D. W. (1973). Technology and decision rules in the theory of investment behavior, *Quarterly Journal of Economics* **87**, pp. 523–543, <http://www.jstor.org/stable/1882023>.
- Levhari, D. and Sheskinski, E. (1972). Lifetime excess burden of a tax, *Journal of Political Economy* **80**, pp. 139–147.
- Loftsgard, L. and Heady, E. (1959). Application of dynamic programming models for optimal farm and home plans, *Journal of Farm Economics* **41**, pp. 51–67.
- Martin, J. and Plaxico, J. (1967). Polyperiod analysis of growth and capital accumulation of farms in the rolling plains of oklahoma and texas, Technical Bulletin 1381, United States Department of Agriculture.
- Moss, C. (1984). *A Polyperiod Analysis of Financial Performance with Alternative Resource Endowments, Credit Constraints, and Organization Form*, Master's thesis, Oklahoma State University.
- Moss, C. (2014). Production economics: An empirical approach, <http://ricardo.ifas.ufl.edu/aeb6184.production/productionbook.pdf>, unpublished Textbook Draft.
- Moss, C. B. (2002). *Handbook of Applied Optimization*, chap. Applied Optimiza-

- tion in Agriculture (Oxford University Press), pp. 957–966.
- Moss, C. B. (2010). *Risk, Uncertainty, and the Agricultural Firm* (World Scientific, Hackensack, NJ).
- Moss, C. B., Featherstone, A. M., and Baker, T. G. (1987). Agricultural assets in an efficient multiperiod portfolio, *Agricultural Finance Review* **47**, pp. 82–94.
- Ramirez, O. A., Moss, C. B., and Boggess, W. G. (1994). Estimation and use of the inverse hyperbolic sine transformation to model nonnormal correlated random variables, *Journal of Applied Statistics* **21**, pp. 289–304, <http://www.tandfonline.com/doi/abs/10.1080/757583872>.
- Swanson, E. (1956). Application of programming analysis to corn belt farms, *Journal of Farm Economics* **38**, pp. 408–419.
- Tu, P. N. (???). *Dynamical Systems: An Introduction with Applications in Economics and Biology*.