DECOMPOSITION OF COMPLEX DECISION PROBLEMS
WITH APPLICATIONS TO
ELECTRICAL POWER SYSTEM PLANNING

by

Edward G. Cazalet

DECISION FOCUS, INCORPORATED
1801 Page Mill Road
Palo Alto, California 94304

May 1970

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald A. Howard

(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David G. Luenberger

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard D. Smallwood

Approved for the University Committee on Graduate Studies:

Lincoln E. Moses

Dean of the Graduate Division

ACKNOWLEDGMENTS

I am indebted to many individuals and organizations for financial support, encouragement, guidance, and valuable insights. Among the organizations, I gratefully acknowledge the financial support provided to me by the National Science Foundation under Grant NSF-GK-16125. Financial support for the early phases of this research was provided by the Decision Analysis Group at Stanford Research Institute as part of an internal research project. Also, the Decision Analysis Group, under contract to the Federal Electricity Commission of Mexico, supported the original research underlying the power system example in this dissertation.

Among the individuals whose efforts I wish to acknowledge, the contributions of Dean Boyd are most appreciated. Dean and I worked together during the summer of 1969 to develop computational methods for analyzing multiproject decision problems. A number of the fundamental results reported in this dissertation are the product of our joint efforts. In addition, Dean provided technical advice and stimulating conversations that directly contributed to this dissertation.

My advisor, Professor Ronald Howard, set the standards of excellence that motivated my work. Professor Howard and Professors David Luenberger and Richard Smallwood read the dissertation and supplied several constructive remarks. I also am indebted to my friends at Stanford and SRI who contributed in many ways. The typing was done by Mrs. Sheila Hill who has clearly lived up to her reputation as an excellent technical typist.

I owe much to my wife, Diane, for her helpful editorial comments. Most importantly, I am grateful to my wife and children for their love and encouragement.

ABSTRACT

In this dissertation a methodology for analyzing complex decision problems is developed and is applied to an electrical power system planning problem. The methodology is based on the idea of decomposing a complex problem into a number of simpler subproblems and then coordinating their solution to solve the original complex problem.

The methodology is primarily designed for strategic decision problems where a computerized model is appropriate. The methodology has two main parts: The first part is concerned with structuring the model for the decision problem, identifying the subproblems, and selecting a method of coordinating the subproblems; The second part provides a mathematical foundation for solving problems by decomposition.

In Chapter 2 the mathematical foundations of decomposition for deterministic decision problems are developed by using a series of increasingly complicated examples. In each example, a decision problem is interpreted as a resource allocation problem among a number of independent projects where the resources must be purchased in a resource market. The allocation of the resources among the projects is usually coordinated through a pricing scheme. By assigning a price to each resource and then adjusting the prices in an organized way, an optimal solution to the original decision problem is obtained. These examples show how to decompose decision problems involving time, multiple resources and multi-attribute preference structures.

The mathematical foundations developed in Chapter 2 are based on a

simple theorem that provides a test for the optimal solution to a resource allocation problem. The theorem provides sufficient conditions for an optimum so that trial solutions that pass the test are guaranteed to be globally optimal. Solutions that do not pass the test may or may not be optimal. The theorem is applicable to nonlinear problems with decision variables defined on either discrete or continuous sets.

In Chapter 3 the first part of the methodology is demonstrated by applying it to a capacity expansion problem for an actual electrical power system. Detailed models of the generating plants, system operating cost, and system reliability are developed. Decomposition of this problem provides several insights into power system planning and shows how to decompose problems with complex technical interactions between projects.

A mathematical foundation for decomposition under uncertainty is developed in Chapter 4. All of the results of Chapter 2 are extended to problems under uncertainty including problems with a multi-attribute risk preference function (von Neumann-Morgensterm utility function). A new notation for describing decision problems under uncertainty plays a key role in this chapter. The results of this chapter can be applied to decision problems where the uncertainty is resolved gradually, or quickly, over time and the decisions are dynamically adjusted in response to new information.

The analysis of the electrical power system is extended in Chapter 5 to include uncertainty in a crucial state variable. The solution of this problem demonstrates the computational feasibility of decomposition under uncertainty. The results of this chapter, together with Chapter 2, demonstrate that every aspect of this difficult power system problem can be treated by the methodology.

The methodology can be applied to any strategic decision problem, although it is most useful in problems with many decision variables. When it is not appropriate to gather all of the information relevant to the problem in one place, then the methodology suggests ways to decentralize the problem so that the decisions are delegated to several decision makers.

This dissertation provides significant contributions to power system planning and to the theory of decomposition under conditions of certainty and uncertainty. The most important contribution, however, is a complete methodology for solving a class of complex decision problems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Strategic planning problems are often characterized by their importance, complexity, dynamic effects, uncertainty, and complex preferences. Many strategic planning problems require computerized models if a careful analysis is to be performed. In this dissertation a methodology is developed for analyzing complex planning problems where detailed models are appropriate.

At present, the analyst's tools are rather limited in situations where complex models are required. Generally, the analyst must choose between the following two approaches to modeling and optimization:

1. Standard modeling and optimization methods such as linear programming which solve an approximation to the actual problem.
2. Detailed simulation models which require heuristic optimization methods.

While some intermediate choices exist, they do not provide the generality required for analyzing many complex problems. In this dissertation very general modeling and optimization methods are developed that take advantage of the natural structure of a complex problem rather than imposing a restrictive structure on the problem.

In this chapter many features of the methodology will be discussed.

First, however, the objectives of the dissertation will be formally stated.

1.1 Statement of Objectives

The dissertation has two primary objectives. They are

1. To develop a methodology for the solution of strategic decision problems where detailed models can be economically justified, and
2. To apply the methodology to electrical power system planning.

The restriction to strategic decision problems implies that the methodology is not designed for analyzing the tactical decisions such as the decisions encountered in the daily operation of an electrical power system. The restriction to computerized models implies that all of the available information relevant to the decision can be gathered in one place, as opposed to the situation in decentralized organizations where most of the detailed information is diffused among several decision makers and experts. Some of the reasons for these restrictions will become clear as we proceed. At many points in the development of the methodology we will indicate which portions of the methodology apply to a more general class of problems.

1.2 Introduction to the Basic Concepts of the Methodology

The methodology is based on the idea of decomposing† complex

† It is important to note that the term "decomposition" is often used to describe the situation where two or more subproblems (or systems) do not interact or where the interactions are insignificant. In such

problems into a number of simpler, independent subproblems and then coordinating their solution to solve the original problem. Coordination of the decomposed subproblems usually is achieved through a pricing scheme. By adjusting the prices in an organized way a solution to the original problem usually can be obtained.

The methodology can be divided into two main parts. The first part is concerned with structuring the model for the decision problem, identifying the subproblems, and selecting a method of coordinating the subproblems. The second part of the methodology provides a mathematical foundation for solving problems by decomposition.

In the first part of the methodology it is often useful to interpret problems in terms of a resource allocation problem among a number of independent projects where the resources are purchased in a resource market. When the prices (marginal costs) of the resources are provided, then it is relatively easy to determine the optimal amount of resources that each project should consume. By adjusting the prices in ways that will be further described, the allocation of resources among the projects can be coordinated so that the optimal allocation for the whole problem is achieved.

The key to devising effective computational methods is to define the resources and projects so that the projects are independent when

---

cases decomposition is easy to achieve. In this dissertation we are generally concerned with subproblems that interact in some way. The decomposition achieved by the methods of this dissertation is not decomposition in the strict sense that the subproblems do not interact. Rather, we say that given the prices on the resources, which cause the interactions, we can act as if the subproblems do not interact. The determination of the appropriate prices is considered to be a separate problem.

when the prices of the resources are provided. One approach to identifying the resources and projects is to use calculus to obtain the necessary conditions for an optimal allocation. The necessary conditions can be expressed as a set of simultaneous equations where the decision variables are the unknown variables. By considering iterative methods for solving these equations, insight can be developed into the problem of defining resources and projects. When calculus cannot be applied to a problem the approach just described still provides insight, but the second part of the methodology must be applied to justify the resulting computational methods.

The second part of the methodology is based on a simple theorem that provides a test for the optimal solution to a resource allocation problem. The theorem provides sufficient conditions for an optimum so that trial solutions that pass the test are guaranteed to be globally optimal. Solutions that do not pass the test may or may not be optimal. Thus, in a sense, the theorem provides a "fail-safe" test for the optimal solution to a resource allocation problem, since it never indicates a trial solution is optimal when a better allocation is possible.

The conditions of the theorem define two optimization problems that are related to the original problem. If the two problems have the same solution, then the theorem guarantees that this solution is the solution to the original problem. Significantly, the theorem is applicable to nonlinear problems where the decision variables are defined on either discrete or continuous sets.

The theorem is initially developed to be applied to relatively simple problems. However, the theorem can be easily extended to apply to very complex problems. The extension of the theorem to more complex problems requires no important new concepts. In its most general form the theorem applies to problems involving time, uncertainty and complex preferences.

In practice, the methodology does not provide an inviolable procedure for analyzing a strategic decision problem. Usually, an analysis of a problem is performed iteratively, in the sense that successive improvements are made in the formulation of the problem and design of the computational methods. In order to illustrate some of these practical aspects of the methodology, it is useful to consider an example. In this dissertation the methodology is applied to the analysis of capacity expansion decisions in an actual power system. The analysis of this power system example will be discussed in detail in Chapter III. The following section provides an informal introduction to the example and insight into the method of analysis.

1.3 Applications to Electrical Power System Planning

An important strategic decision problem in electrical power system planning concerns the installation of new generating plants.

The decisions in this problem include selecting the size, type, and date of installation of the new generating plants.

In Chapter III an actual electrical power system problem is solved by decomposing it so that each alternative generating plant is viewed as a project or subproblem. In this section the results of the example will be summarized in terms of a hypothetical decentralized organization designed specifically to plan and operate the power system. This summary also provides insight into developing organizational interpretations for other complex problems.

At the head of this hypothetical organization is the president who bears the ultimate responsibility for planning the power system. Normally, he does not make the major decisions. Instead, the decisions are delegated to the plant and system managers.

The plant managers are responsible for installing generating plants. For example, one of the plant managers is responsible for installation decisions for one type of plant (hydro, nuclear, conventional thermal, or gas turbine) in one particular year. A plant manager's decisions include choosing the size of the plant and possibly selecting optional equipment and financing methods.

The system managers are concerned with the operation of the system. Two types of system managers are together responsible for meeting the demands for electricity. The first type of system manager is the operating system manager. He is concerned with selecting the best system for generating electricity. The second type of system manager is the reliability system manager, who is concerned with selecting the best system for assuring that the demands for electricity can be met. For each year in the planning period it is useful to hypothesize distinct system managers.

If each of the plant and system managers could act as if each were running an independent business then the problem of planning the expansion of the power system would be relatively easy. Unfortunately, the system managers prefer efficient and reliable plants while the plant managers, who do not bear the costs of operation and service outages, prefer inefficient and unreliable plants because they cost less. Clearly, some form of coordination is necessary.

One way the decisions can be coordinated is to compensate managers for costs they incur because of the actions of other managers. If the mechanism for compensating the managers is carefully designed then each manager will still retain a degree of independence.

The simplest method of compensating the managers is a pricing scheme. A pricing scheme, for example, sets prices on all the services (resources) provided by the plant managers.[†] The system managers are required to compensate the plant managers at these prices for the quantity of each service provided. For the particular electrical system considered in this dissertation the services include the total capacity of each type of plant, the amount of hydro energy available, the hourly operating costs of the plants, and the average available total capacity based on reliability considerations. Separate services and prices

---

† In this dissertation we carefully distinguish between the "price" of a resource and the "cost" of a resource. The cost of a number of units of a resource is the price of the resource times the number of units. Thus, for example, the cost is measured in dollars, whereas the price is in dollars per unit.

on the services are defined for each year in the planning period.

Theoretically, the entire decision problem must be solved to determine the prices of the services. For the optimal policy the price of a service is the value of an additional unit of service provided to the system managers. A practical method is to estimate the prices and then successively adjust the prices until the correct prices are obtained.

There are many methods of successively adjusting the prices. In one of the methods the organization's president initially estimates the prices. Given the prices the managers in the decentralized organization independently choose the amounts of each service that they would provide or consume at these prices. If the managers happen to agree on the amounts of services, then the president has correctly estimated the prices. If the managers do not agree, then the difference between the proposed amount of each service provided and the amount consumed indicates whether the price on that service should be increased or decreased. For example, if more hydro capacity is demanded by the system managers than is supplied by the plant managers, then the price of hydro capacity is too low and should be increased on the next iteration of the prices.

This approach to planning by a decentralized organization is analogous to the decomposition of a detailed model of the planning problem. The decomposition approach is computationally superior to a direct approach if the number of iterations required is small. Usually, the computational effort required for each iteration is orders of magnitudes less than the computational effort required to solve the entire problem directly. Thus, significant overall computational savings are possible with the method described above.

The results of this dissertation can be viewed as providing a theoretical foundation for the intuitive decomposition methods described above. An important part of the methodology developed in this dissertation is directed at the problem of identifying the services or resources that are priced to coordinate the independent subproblems.

1.4 Summary of Results

In Chapter II a theoretical foundation for the solution of extremely general optimization problems is developed. While the theory is formally valid for almost any optimization problem the results of the theory are of practical interest only for unconstrained optimization problems having special types of structure.

The theory is based on a simple mathematical result that provides a test for the optimality of a trial solution to an optimization problem. The test provides sufficient conditions for an optimum so that any trial solution that passes the test is guaranteed to be a global optimum. Trial solutions that do not pass the test may or may not be optimal.

Two general methods for searching for the optimal solution to a problem are developed. Neither of the methods can be guaranteed to converge rapidly for all possible problems. However, the issue of convergence is not crucial for strategic decision problems of the type that require a detailed computer model. For this class of problems, the analyst can afford to interact with the computer to choose the best method for solving a large problem.

The computational power of the theoretical methods depends on the optimization problem having certain special structure. Usually this structure can be interpreted in terms of a resource allocation problem among a number of independent economic units where the resources must be purchased in various markets. Few problems naturally exhibit the required structure. Often, a complex problem must be carefully formulated to obtain this special structure. However, this approach is very effective for many complex problems that cannot be readily solved by any other method.

The initial theoretical results in Chapter II are mathematically similar to certain methods of solving constrained optimization problems. All of the results of this dissertation are stated in an unconstrained form. For simple problems the distinction between constrained and unconstrained problems is often unimportant. For complex problems the superiority of the unconstrained formulation of problems is evidenced by the success of this dissertation in treating problems with very arbitrary objective functions and complex forms of uncertainty.

The results in Chapter IV on decomposition under uncertainty are valid for extremely general problems. Problems where the uncertainty is slowly, or quickly, resolved over time and the decisions are dynamically adjusted in response to new information can be treated by the methods developed in Chapter IV. A significant result of this chapter is that problems under uncertainty can be decomposed using exactly the same techniques as problems under certainty. There are very few non-trivial results on decomposition under uncertainty in the literature.

The theoretical results in this dissertation are interpreted in terms of hypothetical decentralized organizations wherever possible. Since the mathematical foundations of decomposition and decentralization are similar, this dissertation can be viewed as a contribution to the theory of designing decentralized organizations. Conceptually, this theory can be applied within corporations and at all levels of government. However, none of the practical issues concerning decentralization are considered here.

In Chapter III a complicated electrical power system planning problem is posed and solved by decomposition. The problem is based on an analysis of an actual electrical system. The application to electrical power system planning is a convenient way of communicating an approach to problem formulation. By using this approach, decomposition can be applied to very complex problems.

The analysis of the power system problem requires no more mathematical tools than the simplest problem formulated in Chapter II. The level of mathematics requires elementary calculus at most. The only complicating factor is the notational problem caused by the very size and complexity of the power system problem.

The analysis of the power system problem yields important general insights into power system planning. In Chapter V the analysis is extended to include uncertainty in some of the crucial variables of the power system problem. Thus, this dissertation is both a contribution to electrical power system planning and a methodology for solving complex decision problems by decomposition.

1.5 Related Literature

Many of the basic ideas behind decomposition have been in existence for some time. There is a considerable literature on decomposition and decentralization in the fields of economics, business and operations research.

The results on the decomposition of unconstrained optimization problems in this dissertation have drawn on the vast work on the decomposition of constrained optimization problems. From our point of view, the paper by Everett [12] is the best introduction to the decomposition of constrained optimization problems. The work summarized in Arrow and Hurwicz [2] is particularly relevant to the design of algorithms. Lasdon [20] and Geoffrion [15] have developed other logical aspects of decomposition that are relevant here.

Decomposition of constrained optimization problems under uncertainty is a difficult problem and the literature on the subject is small. One excellent attempt related to Everett's work on constrained problems is in Mitchell [21]. A different approach that uses dynamic and linear programming for problems under uncertainty is by Wilson [29].

The present work is a continuation of a joint research project by the author and D. W. Boyd on decentralization of resource allocation problems. The results of that research project are reported in Boyd and Cazalet [6][7]. The dissertation of Boyd [5] develops a methodology that uses decomposition to assist in the assessment of complex preferences in decision problems that do not have a direct means for economic valuation.

The dissertation of Helms [17] develops an approach to decomposition of unconstrained problems from a different point of view. However, Helms does not explicitly treat uncertainty in his work.

The literature on electrical power system planning is of varying quality. The state-of-the-art is essentially summarized in Nelson [23], Berrie [4], and Turvey [28]. The power system problem examined in this dissertation is based on an analysis performed by the author and his colleagues in the Decision Analysis Group at Stanford Research Institute [10][13]. The analysis was done for the Comisión Federal de Electricidad and it concerned the capacity expansion of the Mexican electrical system with particular emphasis on nuclear power plants.

1.6 Contributions to Decision Analysis

Strategic planning problems provide a unique challenge to management scientists. Decision analysis[†] is one discipline that has addressed itself to the logical solution of complex problems where significant resources are involved. It has well developed quantitative tools for the analysis of the one-shot, single project types of decisions. However, in the area of multiple project, repetitive decisions,[‡] satisfactory computational methods have not yet been developed.

In other disciplines, the powerful techniques provided by mathematical programming have proved to be popular. Linear programming, for example, provides an extremely powerful tool for the analysis of problems that can be structured within its assumptions. But, decision analysts generally have been unable to put the techniques of mathematical programming to work. One of the difficulties is that mathematical

† Introductions to decision analysis are in Howard [18], North [29] and Raiffa [26].

‡ For a definition of multi-project selection decisions and a comprehensive review and critical analysis of the literature on the subject, see Boyd and Matheson [8].

programming under uncertainty is still not fully developed. Another reason is that decision analysts typically do not formulate problems in terms of the constrained models that mathematical programming specifically addresses itself to.

The results of this dissertation are useful in the decision analysis of multiple project, repetitive decisions. The powerful concepts of iteration and decomposition inherent in most mathematical programming techniques are brought to bear on these difficult decision problems. The new computational methods and ways of structuring problems apply to decision problems that are best solved on a centralized basis.

Often the differences between the methods developed in this dissertation and the methods of mathematical programming and elementary calculus are subtle. In the simplest problems the differences between constrained and unconstrained formulations of decision problems are often a matter of philosophy and only rarely do they significantly affect the difficulty of an analysis. In the more interesting problems, where nonlinearity, uncertainty, dynamic effects, and complex preferences are present, the advantages of the methods discussed in this dissertation become evident.

Electrical power system planning is an example of a multiple project, repetitive decision problem. In this dissertation we develop decomposition techniques for a general class of problems and then apply the techniques to electrical power system planning. We demonstrate that every important aspect of this problem can be treated in a practical way. This successful application is strong evidence that the methods developed in this dissertation can provide practical tools for the decision analysis of multiple project, repetitive decision problems.

CHAPTER II

MATHEMATICAL FOUNDATIONS OF DECOMPOSITION

In this chapter we develop the mathematical foundations of a methodology for solving unconstrained optimization problems by decomposition. The mathematical foundations are developed using a series of increasingly complicated examples. These examples do not apply directly to any particular problem; rather, they are suggestive of ways to structure actual problems.

The examples focus on the resources allocated in a decision problem. The first example treats problems with a single resource to be allocated among a number of projects. It is assumed that the resources are purchased in a market where the price of the resource is a function of the amount purchased. This formulation of the problem should be contrasted with the more usual constrained formulation which limits the amount of resources available.

The second example in this chapter extends the results of the first example to problems with multiple resources. The third example treats problems with very general objective functions. Both the second and third examples are applicable to problems over time. The development of the mathematical foundations for decomposition under uncertainty is postponed until Chapter IV.

For each of the examples we prove an optimality theorem, derive bounds on the optimal solution and present two search algorithms. Essentially, all of the basic ideas are introduced in the discussion of the first example.

2.1 Single Resource Problems with Separable Objective Functions

The mathematical ideas developed in this section apply to the following class of problems:

1. The relationship between a given resource allocation and the eventual outcome is known with certainty.
2. The objective of the problem can be interpreted as maximizing profit where profit is separable into a term representing the total project returns and a term representing the cost of a single resource.
3. The cost term depends only on the total amount of a single resource allocated among the projects.

The resources, revenues, costs, and profits should be flexibly interpreted. For example, a resource can be a service, a commodity, or something less tangible.

The Example

Consider the problem of allocating a single resource among J projects. Let

$$x_j \equiv^{\dagger} \text{amount}^{\ddagger} \text{ of the resource used by the } j^{th} \text{ project,}$$

$$j = 1, \ldots, J .$$

$$\underline{x} \equiv (x_1, \ldots, x_J), \quad \text{a vector.}$$

† The symbol "≡" is read "is defined as."

‡ A negative value for x_j indicates a net production of the resource.

$R(\underline{x}) \equiv$ total revenue from all projects as a function of the amount of the resource employed by each project. In general,

$$R(\underline{x}) = \sum_{j=1}^{J} r_j(\underline{x})$$

where $r_j(\underline{x})$ is the return assigned to the j^{th} project.†

$y \equiv$ total amount of the resource used by all J projects. Thus

$$y = \sum_{j=1}^{J} x_j \ .$$

$C(y) \equiv$ total cost of the resource purchased in the resource market.

The resource allocation problem is to choose an $\underline{x}$ to maximize the profit function

$$R(\underline{x}) - C(y)$$

where $\underline{x}$ is chosen from the completely arbitrary set X. The set X can restrict the allocation $\underline{x}$ in any way. For example, the set X can restrict the resource to be available only in discrete units. Later in this section we will require both $R(\underline{x})$ and X to have certain properties.

† At this point, the revenue assigned to the j^{th} project can depend on the allocations to every project. Later we will make the assumption that the revenue assigned to the j^{th} project is independent of the revenue assigned to the other projects. This assumption is not required now.

Many optimization techniques are applicable to the type of problem described above. The techniques range from an exhaustive search over all elements of the space X to sophisticated nonlinear programming techniques. The more powerful techniques utilize certain special characteristics of a problem to guide a search for the optimum and to guarantee that the result is the optimal resource allocation.

When the number of projects in the example resource allocation problem is large, the problem is more difficult to solve. For example, a problem with only 10 possible resource allocations to each project results in an overall problem with 10^J possible resource allocations. In this section we develop methods for solving problems with large numbers of projects.

Statement and Proof of Theorem I

There are two important tasks in the design of optimization methods. One task is to develop methods for recognizing the optimal solution once it is found. The other task is to design efficient methods for finding the optimal allocation. The following theorem relates to the first task. The theorem is useful because it provides a method for testing whether a given resource allocation is optimum.

THEOREM I:[†] If $\underline{x}^*$ maximizes

$$R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j$$

[†] Some of the history behind this theorem is reviewed in the discussion of constrained and unconstrained problem formulations later in this section.

over all $\underline{x} \in X$, and if y^* maximizes

$$\lambda y - C(y)$$

over all y, and if

$$\sum_{j=1}^{J} x_j^* = y^*$$

then $\underline{x}^*$ maximizes

$$R(\underline{x}) - C(y)$$

over all $\underline{x} \in X$.†

Proof:

a) The theorem statement implies the following two inequalities:

$$R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j \leq R(\underline{x}^*) - \lambda \sum_{j=1}^{J} x_j^* \tag{1}$$

holds for all $\underline{x} \in X$, and

$$\lambda y - C(y) \leq \lambda y^* - C(y^*) \tag{2}$$

holds for all y.

b) Combining inequalities (1) and (2) gives

$$R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j + \lambda y - C(y) \leq R(\underline{x}^*) - \lambda \sum_{j=1}^{J} x_j^* + \lambda y^* - C(y^*) \tag{3}$$

which holds for all $\underline{x} \in X$ and all y.

c) Since (3) holds for all y it must also holds for $y = \sum_{j=1}^{J} x_j$

† The statement $\underline{x} \in X$ is read "$\underline{x}$ is an element of the set X."

where $\underline{x} \in X$. In this case the terms involving λ on the left side of (3) cancel, and

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right) \leq R(\underline{x}^*) - C(y^*) + \lambda\left[y^* - \sum_{j=1}^{J} x_j^*\right] \tag{4}$$

holds for all $\underline{x} \in X$.

d) By the statement of the theorem

$$y^* = \sum_{j=1}^{J} x_j^* .$$

Thus, the terms involving λ on the right side of (4) cancel and

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right) \leq R(\underline{x}^*) - C\left(\sum_{j=1}^{J} x_j^*\right) \tag{5}$$

holds for all $\underline{x} \in X$. Hence the theorem is proved.

Discussion of Theorem I

Theorem I can be viewed as a test to be applied to a trial resource allocation. If the trial resource allocation simultaneously maximizes both $R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j$ over all $\underline{x} \in X$ and $\lambda y - C(y)$ over all y, then the trial allocation is guaranteed to be the globally optimal allocation. However, if the trial allocation fails the test provided by the theorem then we cannot say definitely that the trial solution is not optimum. Stated differently, Theorem I provides only sufficient conditions for an optimum. Thus, in a sense Theorem I provides a "fail-safe" test since it never indicates a trial allocation is optimum when there is a better allocation possible.

The functions $R(\underline{x})$ and $C(y)$ are completely arbitrary (except they must be real-valued). The set X is also completely arbitrary. Thus, Theorem I is valid for an extremely large class of problems including discrete, nonlinear problems.

We can get a considerable amount of insight into Theorem I by studying its application graphically. An application of the theorem is illustrated in Figure 2.1. The horizontal axis of the figure is the total consumption of the resource by all J projects. The vertical axis is the total revenue from all J projects. Each of the points in Figure 2.1 represents a particular allocation of resources.

The resource cost function is also plotted in Figure 2.1. The vertical axis of the figure represents total cost when we refer to the resource cost function. Since the profit obtained from a given policy is just revenue less cost, the vertical distance between a policy and the resource cost function in Figure 2.1 is the overall profit of that policy. The resource cost function shown in Figure 2.1 happens to be convex.†

The maximization of $R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j$ required in Theorem I can be interpreted graphically. Consider a hyperplane of slope λ. In Figure 2.1 this hyperplane is a straight line of slope λ. Now, lower this hyperplane until it touches a policy in the revenue-resource space. The first policy touched by the hyperplane of slope λ, maximizes $R(\underline{x}) - \lambda y$, where $y = \sum_{j=1}^{J} x_j$. Call this policy $\underline{x}^*$.

† A function is convex if the function always lies below or on a line drawn between any two points on the graph of the function, i.e., $c(y) \leq \alpha c(y_1) + (1-\alpha)c(y_2)$ where $0 \leq \alpha \leq 1$. A function is strictly convex if the inequality holds in the strict sense. A function is (strictly) concave if the negative of the function is (strictly) convex.

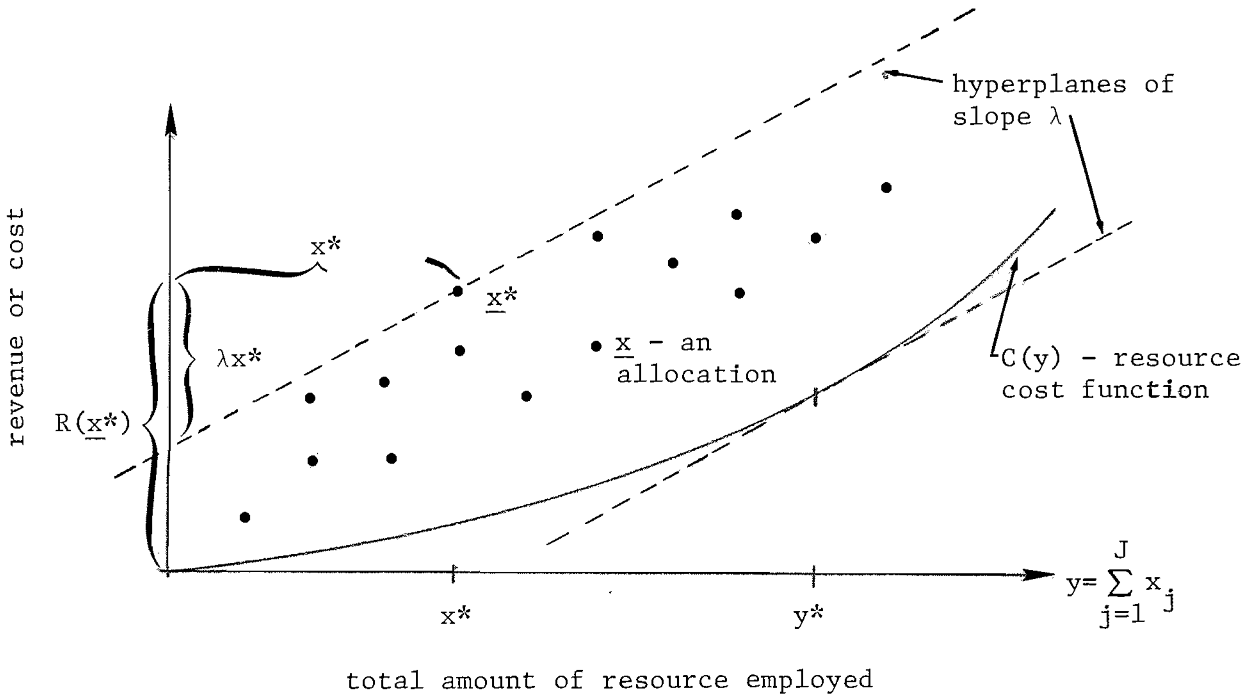


Figure 2.1: AN EXAMPLE WHERE THE CONDITIONS OF THEOREM I ARE NOT SATISFIED

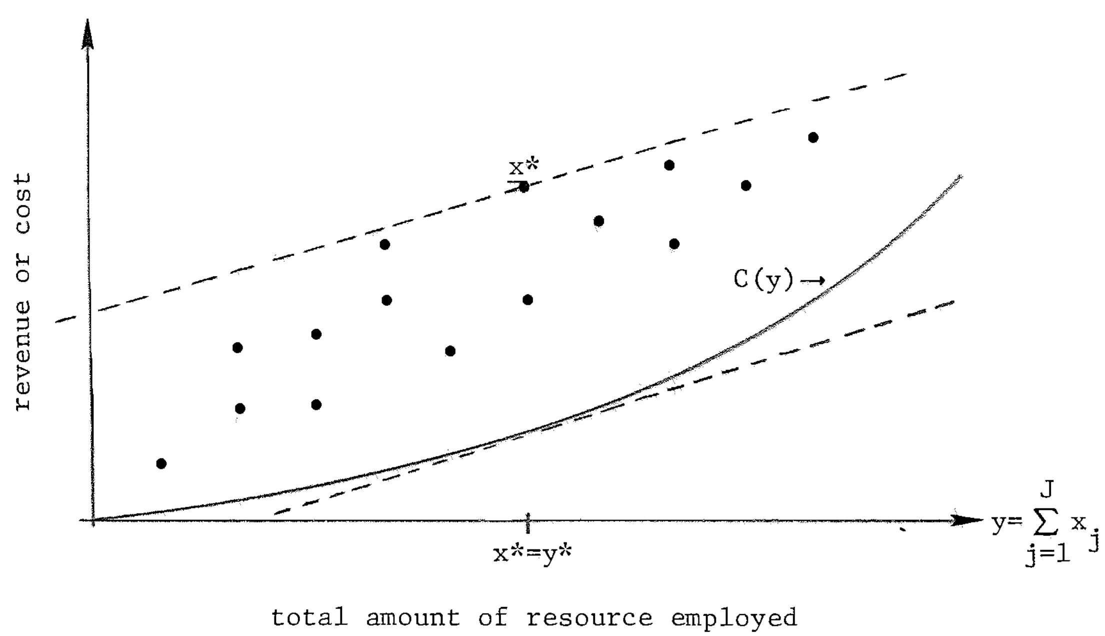


Figure 2.2: AN EXAMPLE WHERE THE CONDITIONS OF THEOREM I ARE SATISFIED

By a similar operation we can determine the total amount of the resource required to maximize $\lambda y - C(y)$, or, equivalently, to minimize $C(y) - \lambda y$. To minimize $C(y) - \lambda y$ we raise a hyperplane of slope λ until it touches a point on the resource cost function. The amount of resources required at that point is the amount that maximizes $\lambda y - C(y)$.

The statement of Theorem I says that if the $\underline{x}^*$ that maximizes $R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j$ also maximizes $\lambda y - C(y)$ where $y = \sum_{j=1}^{J} x_j$ then $\underline{x}^*$ is the optimal allocation. In Figure 2.1 the conditions of the theorem are not satisfied because the parallel hyperplanes do not generate the same total resource allocation. However, in Figure 2.2 the conditions of the theorem are obviously satisfied. The key to a successful application of Theorem I is to choose the correct value for λ. We will discuss a number of methods for adjusting λ later in this section.

Several graphical applications of Theorem I are presented in Figure 2.3. Some of the examples illustrate cases where the theorem guarantees an optimum if the correct λ can be found. The other examples illustrate cases where the theorem cannot guarantee that any of the solutions is the optimal solution, regardless of the slope of the hyperplane employed.

All of the examples in Figure 2.3 illustrate the application of Theorem I to problems with continuous revenue and cost functions. The total revenue is defined in each case as

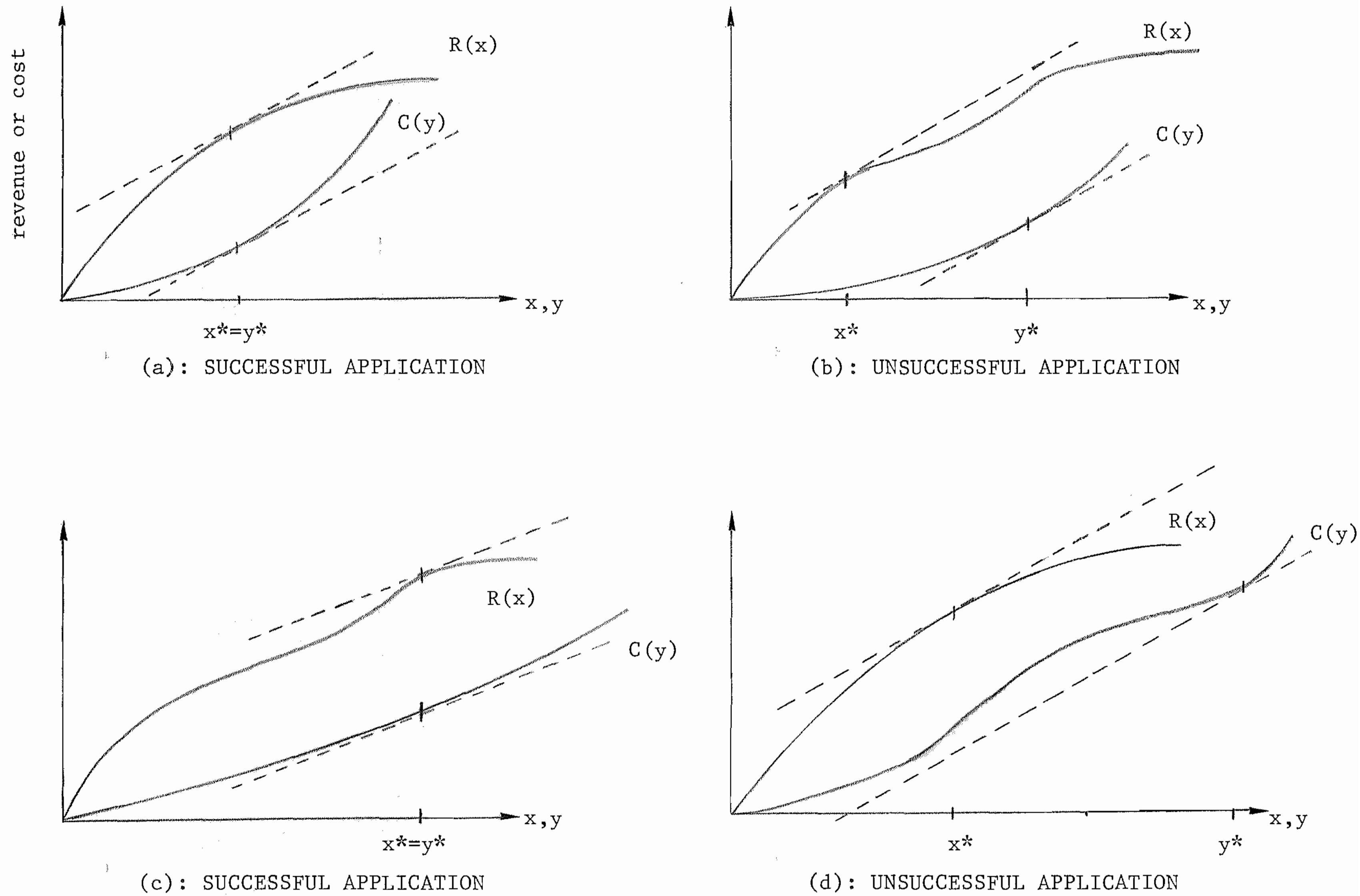


Figure 2.3: FURTHER EXAMPLES OF THE APPLICATION OF THEOREM I

$$R(x) = \max_{\underline{x} \in X} R(\underline{x}) \quad \text{subject to} \quad \sum_{j=1}^{J} x_j = x$$

so that we can use a two-dimensional presentation.

Example a) in Figure 2.3 illustrates a case where the total revenue function is concave and the cost function is convex. The solution determined by the parallel hyperplanes satisfies the conditions of Theorem I.

Example b) illustrates a case where the theorem cannot guarantee the optimality of a policy even though an optimal policy obviously exists. The difficulty arises because the optimal policy lies in a "gap."[†] When the maximum of $R(x)$ is determined by lowering a hyperplane it is not possible for the hyperplane to reach into the gap in the total revenue function.

Gaps in the total revenue function exist only when the total revenue function is not concave. Nevertheless, Theorem I is still useful in problems with non-concave total revenue functions, if the optimal policy is not in a gap. Example c) illustrates a case where the theorem is able to guarantee the optimality of a policy even though $R(x)$ is non-concave.

Example d) in Figure 2.3 demonstrates that gaps also can exist in the cost function if it is non-concave. Again, Theorem I will guarantee the optimality of a policy only if the policy does not lie

[†] A gap is a well-defined term in the literature on mathematical programming. A paper by Everett [12] popularized the term. We will discuss methods of resolving gaps in the subsection on penalty functions.

in a gap. When the optimal policy lies in a gap Theorem I simply has nothing to say. In such cases the analyst is forced to use other techniques including the extensions of Theorem I discussed later in this section.

Constrained and Unconstrained Problem Formulations

The relationship between constrained and unconstrained formulations of a problem is central to the problem of developing a methodology for formulating and solving complex problems. We must, however, be careful in our discussion of this topic, because the adjectives "unconstrained" and "constrained" never completely describe a given problem formulation.

In discussing constrained and unconstrained problem formulations there are two important points to consider.

The first point concerns how well a model describes the decision maker's view of his problem. If there is an overwhelming physical or economic reason why a particular variable in a problem should be restricted then a constraint on that variable is a good modeling approximation. On the other hand, if a variable is constrained for analytical reasons then it is important to test the sensitivity of the ultimate decision to the level of the constraint. Lagrange multipliers are particularly useful in this regard.

The second point concerns the difficulty of the analytical problem. A constrained problem formulation eliminates the need for detailed modeling of certain features of the problem. In the simple example used in this section a constraint on the amount of the resource available would eliminate the need for a model of the resource market. Another apparent advantage of constrained problems is that mathematical programming techniques for constrained problems are fairly well developed.

The advantages of a generally unconstrained formulation of a problem are particularly important in strategic decision problems. In strategic decision problems the only realistic constraints are "physical constraints." Physical constraints, for example, are the number of hours in a day or the availability of a resource in discrete amounts.

Constrained formultions of problems tend to be least realistic when time, uncertainty and multiple outcomes must be explicitly treated. Very often an analyst will attempt to avoid the hard analysis required to construct a preference model for these situations. Instead he will use constraints to eliminate some outcomes from consideration. The result is that a clear understanding of the preferences of the decision maker is avoided, but often at the expense of not satisfying the decision maker.

Advances in the theory of preference models and more experience in the construction of resource cost models will make the use of unconstrained models of decision problems less difficult. One of the objectives of this dissertation is to develop additional optimization methods for unconstrained problems.

At the theoretical level, optimization techniques for constrained and unconstrained problems are strongly related. A good example is the classical method of Lagrange multipliers which transforms a constrained optimization problem into a series of unconstrained problems. This approach is particularly well-expressed in Everett [12] where he illustrates how a hard problem can sometimes be decomposed by using Lagrange multipliers.

It is interesting to compare Everett's main theorem with Theorem I. Consider the following constrained optimization problem which is mathematically identical to our single resource market problem:

$$\max_{\underline{x}\epsilon X, y} \; R(\underline{x}) - C(y)$$

$$\text{subject to} \quad \sum_{j=1}^{J} x_j - y = 0 \; .$$

The constraint can be eliminated by defining a single Lagrange multiplier and formulating the Lagrangian as follows:

$$L(\underline{x}, y, \lambda) = R(\underline{x}) - C(y) - \lambda \left[\sum_{j=1}^{J} x_j - y \right] .$$

For this problem, Everett's main theorem states: if, for some value of λ, the quantities $\underline{x}^*$ and y^* maximize $L(\underline{x}, y, \lambda)$, and if $\sum_{j=1}^{J} x_j - y = 0$, then $\underline{x}^*$, y^* is the optimal allocation. Clearly, Theorem I is a restatement of Everett's main theorem.[†]

The essential difference between Everett's theorem and Theorem I is that Theorem I is expressed in a form that is more natural to unconstrained problems than is Everett's theorem. This rather subtle difference between the two theorems will become crucial when we extend Theorem I to more complicated situations. The unconstrained approach embodied in Theorem I will allow us, in later sections, to treat problems with multiple objectives, time dependence, uncertainty and complex technical

---

[†] Actually Everett's theorem is more general since it allows inequality constraints at the expense of requiring the Lagrange multiplier to be positive. In the present context, only equality constraints are relevant. Thus negative Lagrange multipliers are permissible here.

interactions between projects. These problems have not been solved satisfactorily by constrained methods.

In this dissertation we sometimes find it useful to study constrained problems for the insight they provide and most importantly for the mathematical techniques that have been developed for constrained optimization problems. Some of the algorithms and most of the basic mathematical tools used in this dissertation were originally developed for constrained problems. Another reason for studying constrained problems is that there may be portions of some strategic decision problems that involve physical constraints.

Search Methods

As mentioned earlier, the first task in the design of optimization methods is to develop an optimality test. For our simple example, Theorem I provides such a test. The second task is to design efficient methods for finding the optimal solution. Generally a search method is expressed in the form of an algorithm.

An algorithm is a detailed set of instructions for moving from one solution to another with the objective of quickly converging on the optimal solution. In this dissertation we take a flexible approach towards the design of algorithms. We will outline a number of algorithms and discuss their specific features. In practice, however, the analyst will normally consider a variety of algorithms to solve a given problem. More than one algorithm may be used in the solution of a single problem. Fortunately, the optimality test provided by Theorem I is independent of the method used to find the optimal solution.

Bounds

Upper and lower bounds on the optimal profit of a resource allocation problem are useful aids in searching for the optimal allocation. If the upper and lower bounds are sufficiently close at a particular stage in a search then the search can be discontinued and the present solution can be taken as optimal "for all practical purposes."

Bounds are also useful in another way. If a solution lies in a gap, then Theorem I cannot guarantee the optimality of the solution. However, if a solution can be obtained that is sufficiently close to the upper bound on profit, then there is no need to probe the gap for a better solution.

The method of bounding the profit is illustrated in Figure 2.4. Let $\underline{x}'$ be the resource allocation that maximizes

$$R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j$$

over all $\underline{x} \in X$. Then a lower bound on the optimal profit is given by

$$p^{\ell} = R(\underline{x}') - C\left(\sum_{j=1}^{J} x_j'\right)$$

which is just the profit resulting from the allocation $\underline{x}'$.

Now, let y' be the amount of resources required to maximize

$$\lambda y - C(y)$$

over all y. An upper bound on the optimal profit is given by

$$p^{u} = R(\underline{x}') - C(y') + \lambda\left[y' - \sum_{j=1}^{J} x_j'\right] .$$

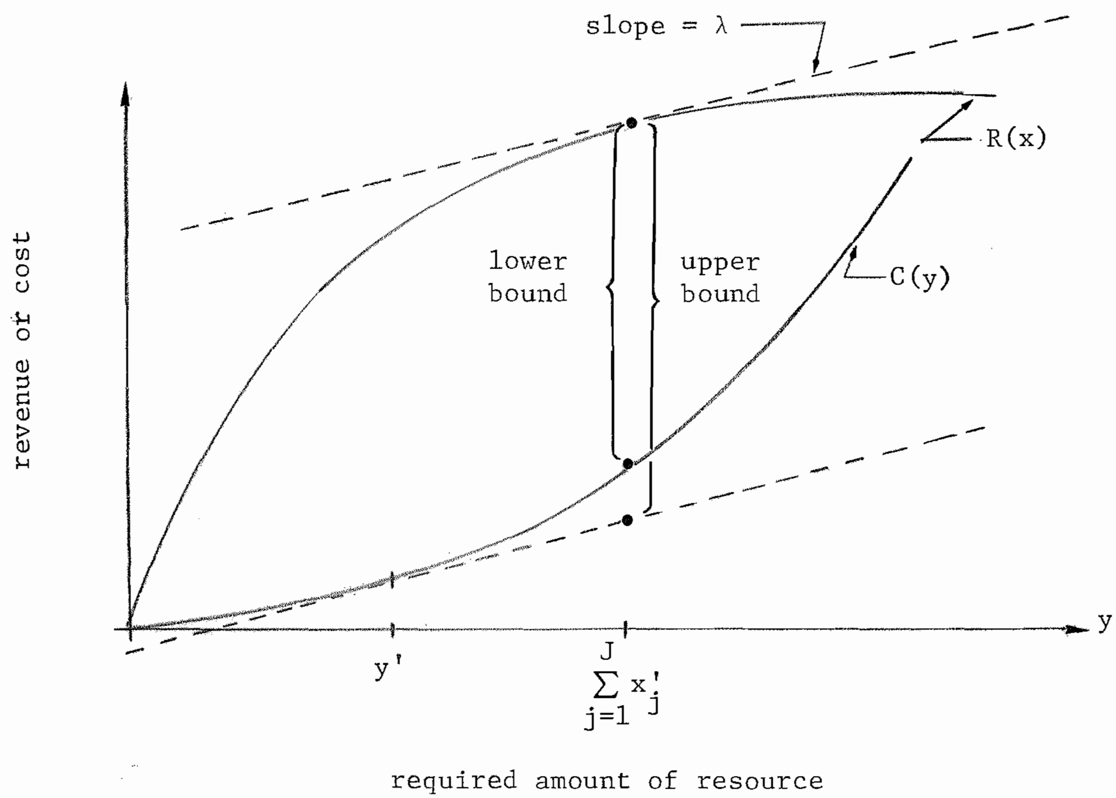


Figure 2.4: BOUNDS ON OPTIMAL PROFIT

At each stage in a search the relevant lower bound is the greatest lower bound obtained so far. Similarly, the relevant upper bound is the smallest upper bound obtained so far. At the optimal allocation, $p^{\ell} = p^{u}$.

The proof that p^{u} is an upper bound follows directly from the proof of Theorem I. Inequality (4) in the proof of the theorem can be rewritten in the form,

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right) \leq R(\underline{x}') - C(y') + \lambda\left[y' - \sum_{j=1}^{J} x_j'\right]$$

which holds for all $\underline{x} \in X$. The term on the right side is the upper bound on the optimal profit.

In some problems it is possible to obtain tighter bounds that depend on the particular structure of the problem. One such alternative method is developed in Boyd and Cazalet [6].

Successive Approximations Algorithm

The first formal algorithm we will investigate is called the successive approximations algorithm [6]. The algorithm is named for its similarity to the method of solving sets of equations by the classical method of successive approximations.

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial λ, λ^0 or start with a trial total allocation at Step 3.

2. Maximize

$$R(\underline{x}) - \lambda^n \sum_{j=1}^{J} x_j$$

over all $\underline{x} \in X$. Call the result $\underline{x}^n$.†

3. Calculate λ^{n+1} according to the relationship

$$\lambda^{n+1} = \frac{d}{dy} C(y) \Bigg|_{y = \sum_{j=1}^{J} x_j^n} .$$

$C(y)$ must be convex in this algorithm.

4. If λ^{n+1} equals λ^n then the conditions of Theorem I are satisfied, and $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Otherwise, return to Step 2 using λ^{n+1}.

The successive approximations algorithm requires that $C(y)$ be convex and differentiable. Under these assumptions Step 3 is equivalent to finding a λ such that $\lambda y - C(y)$ is maximized over all y at $y = \sum_{j=1}^{J} x_j^n$. Steps 2 and 3 and the condition in Step 4 that successive λ's be equal, combine to satisfy the conditions of Theorem I.

Bounds on the optimal profit can be computed at each stage of the successive approximations algorithm.‡ In terms of the notation used to define bounds in the previous subsection, λ is equivalent

† The superscripts on λ^n and x^n are an index to the number of iterations. We could have used the rather cumbersome notation $\lambda^{(n)}$, for example. However, the distinction between $\lambda^{(n)}$ and λ raised to the n^{th} power will always be clear from the context of the application.

‡ A relaxation version of the successive approximations algorithm is described later in this subsection. With a relaxation coefficient not equal to unity, the calculation of an upper bound is not as direct.

to λ^n, $\underline{x}'$ is equivalent to $\underline{x}^n$, and y' is equivalent to $\sum_{j=1}^{J} x_j^{n-1}$. An upper bound on profit cannot be computed until the second iteration of the successive approximations algorithm. The bounds require essentially no additional computational effort.

The statement of the successive approximations algorithm provides an economic interpretation of the term λ. In Step 3, λ is given by

$$\lambda = \frac{d}{dy} C(y) \bigg|_{y = \sum_{j=1}^{J} x_j^n} .$$

In this case, λ is the marginal cost of the resource at the operating point $y = \sum_{j=1}^{J} x_j^n$. If the project revenue functions are differentiable then, at the optimum, marginal revenue is equal to marginal cost.

The economic interpretation of λ as a marginal cost reveals the connection between the methods discussed here and the methods of marginal cost pricing that have long been popular in economics [23], [28]. However, the methods discussed here do not require the revenue functions to be differentiable. In many of the algorithms that we shall propose the cost function $C(y)$ need not be differentiable. Hence, we will usually refer to λ as a "price" rather than a marginal cost. Further economic interpretations are discussed later in this section.

We will not study the convergence of algorithms in detail. This does not imply that convergence is always easy to obtain in the algorithms that we will use. More theoretical study of the algorithms would have practical value in that it provides insight for designing algorithms.

Fortunately, in the class of strategic decision problems treated in this dissertation it is relatively easy for the analyst to interact with the computer to facilitate fast convergence. Thus, the analyst can modify algorithms or choose a new algorithm as the problem is being solved.

The mathematical study of convergence generally requires overly strict assumptions that, in practice, are not always required for fast convergence. In order to say something in general about convergence of an algorithm it is usually necessary to make some statements about the continuity of the functions. Since many practical examples involve discrete functions, the study of convergence might provide insight for discrete problems but would not be directly applicable. The convergence of the successive approximations algorithm is considered in Boyd and Cazalet [6] and again in Boyd [5].

One approach towards improving convergence of the successive approximations algorithm is to introduce a "relaxation coefficient." Let λ^n be the price of the resource determined on the previous iteration. The new price λ^{n+1} is given by

$$\lambda^{n+1} = \alpha \frac{d}{dy} C(y) \Bigg|_{y = \sum_{j=1}^{J} x_j^n} + (1-\alpha)\lambda^n$$

where $0 \leq \alpha \leq 1$.

This calculation replaces Step 3 of the successive approximations algorithm. For α equal to unity the algorithm is as before. For α less than unity the algorithm will converge more quickly in situations where successive solutions tend to oscillate about the optimal solution.

A disadvantage in using a relaxation coefficient in the successive approximations algorithm is that an upper bound is more difficult to determine. Practically, one must first perform Step 2 of the next iteration with α equal to unity in order to compute the current bound and then repeat Step 2 with α less than unity.

Price Directive Gradient Algorithm

Our second formal algorithm is in some ways more sophisticated than the successive approximations algorithm. The price directive gradient algorithm attempts to minimize the upper bound on profit by intelligent choices of successive values for λ.

We have already shown that

$$p^u = R(\underline{x}') - C(y') + \lambda'\left[y' - \sum_{j=1}^{J} x_j'\right]$$

provides an upper bound on the optimal profit. The rate of change (gradient) of the upper bound with respect to λ is given by

$$\frac{\partial p^u}{\partial \lambda} = y' - \sum_{j=1}^{J} x_j' .$$

Economically, the result says that the rate of change of the upper bound is just equal to the excess supply of the resource.

If we view our resource allocation problem as the problem of minimizing the upper bound p^u then it is reasonable to move λ in the direction of the gradient of p^u with respect to λ. This observation suggests the following rule:

1. If the excess supply is positive, then decrease λ.
2. If the excess supply is negative, then increase λ.

We can state this rule algebraically as follows:†

$$\lambda^{n+1} = \lambda^{n} - \alpha\left[y^{n} - \sum_{j=1}^{J} x_j^{n}\right]$$

where

α $\equiv$ an appropriately chosen positive constant.

λ^{n} $\equiv$ price of resource on the n^{th} iteration.

λ^{n+1} $\equiv$ price of resource on $n + 1^{st}$ iteration.

y^{n} $\equiv$ y' on the n^{th} iteration.

x_j^{n} $\equiv$ x_j' on the n^{th} iteration.

One way of choosing the constant α is to use the α that minimizes p^{u} in the direction of the gradient. In practice, however, the analyst would usually adjust α on the basis of a number of intuitive inputs. In multiple resource market problems the direction information provided by the gradient is particularly valuable.

The price directive gradient algorithm can be summarized as follows:

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial price λ^{0}.
2. Maximize

$$R(\underline{x}) - \lambda^{n} \sum_{j=1}^{J} x_j$$

over all $\underline{x} \in X$. Call the result $\underline{x}^{n}$.

3. Maximize

$$\lambda y - C(y)$$

over all y. Call the result y^{n}.

† The symbol α used in this subsection is not to be confused with the relaxation coefficient described in the previous subsection.

4. If $\sum_{j=1}^{J} x_j^n = y^n$, then the conditions of Theorem I are satisfied and $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Otherwise, compute a new value of λ according to

$$\lambda^{n+1} = \lambda^n + \alpha\left[y^n - \sum_{j=1}^{J} x_j^n\right]$$

and return to Step 2.

Convergence of the price directive algorithm depends on the constant α. For small values of α, the convergence of the algorithm tends to be slower but more certain. In problems where convergence is difficult to achieve, a smaller α will often provide convergence.

Mathematically, this algorithm is identical to the price directive algorithm for constrained problems. The name of the algorithm follows Geoffrion [15]. The algorithm is extensively studied in Arrow and Hurwicz [2], and Lasdon [20]. They show that the algorithm will converge for sufficiently small values of the constant α, if the revenue function is concave and the cost function is convex and either of these conditions hold in the strict sense.

There is, however, an important difference between the unconstrained version of the algorithm presented here and the constrained version studied in the references. In the constrained version of the algorithm it is mathematically impossible to implement the intermediate solutions obtained by the algorithm. In a constrained formulation of a problem, only the optimal solution is (primal) feasible.

In the unconstrained version of the price directive algorithm intermediate solutions can always be implemented because the resource market can absorb the excess supply or demand for the resource. Furthermore, the intermediate solutions provide a lower bound on the optimal profit that is not available in the constrained version. The importance of the differences between these two versions of this algorithm will become clearer in later sections of this dissertation.

Decomposition

The computational costs associated with the proposed algorithms depend on the difficulty of the optimization problems that are imbedded in the algorithms. The imbedded optimization problems can be as difficult as the original problem when completely arbitrary project revenue and resource cost functions are assumed. The computational advantages of the methods proposed in this section arise only when the problem has certain special structure. In many cases this special structure is so valuable that it is worthwhile to reformulate a model to obtain the computational advantages associated with the special structure.

In Step 2 of each of the algorithms the following optimization problem must be solved:

$$\underset{\underline{x} \in X}{\text{maximize}} \; R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j \; .$$

The solution of this problem requires, in general, a J-dimensional search. Thus, this problem is nearly as difficult as the original problem. Furthermore, this problem usually needs to be solved a number of times before the optimal allocation is determined.

Now, consider the case where the revenue from each project is independent of the amount of resource allocated to all other projects. With such independent projects,

$$R(\underline{x}) = \sum_{j=1}^{J} r_j(x_j)$$

where

$r_j(x_j) \equiv$ revenue from j^{th} project as a function of the amount of resource allocated to that project.

Furthermore, suppose that the set X describing the alternatives can be partitioned into the product set formed by

$$X = X_1 \times X_2 \times \cdots \times X_{J-1} \times X_J$$

where $x_j \in X_j$ and $x_j \notin X_k$ for $k \neq j$. This second assumption implies that the allocation of resources to the j^{th} project in no way affects the alternatives available to the k^{th} project when $k \neq j$.

By employing the two assumptions stated above, the J-dimensional optimization problem in Step 2 of the algorithms becomes J one-dimensional problems, i.e.,

$$\max_{\underline{x} \in X} R(\underline{x}) - \lambda \sum_{j=1}^{J} x_j = \sum_{j=1}^{J} [\max_{x_j \in X_j} r(x_j) - \lambda x_j] .$$

Thus, we say the problem decomposes when a price is defined on the resource and the projects are completely independent except for the interactions caused by the resource market.

The computational advantages of decomposition can be very important in large problems. Since the difficulty of a search increases approximately exponentially with the dimensionality of the problem it is often much easier to solve many single-dimensional problems several times rather than solve one, multi-dimensional problem.

In problems where the projects are not obviously independent, a useful approach is to try to identify the cause of the interrelationship. Many times, the interactions between projects can be modeled, at least approximately, in terms of a common resource and a market for that resource. The electrical power system problem in Chapter III is a good illustration of decomposition methods in problems with complex interactions.

Organizational Interpretation of Decomposition

Decomposition is a computational tool for solving decision problems that can be treated on a centralized basis. For example, a company in which all decisions are made by a single individual or group would have problems of this type.

A decentralized organization is one where the decisions are delegated to many individuals or groups. Generally the independent decisions made by this organization must be coordinated in some way.

In this section we will interpret our results on decomposition in terms of a hypothetical decentralized organization. The interpretation provides both insight into decomposition and concepts that are useful in the design of decentralized organizations.†

The basic decentralized organizational structure used in this

† Decentralization is discussed extensively in Arrow [2], Boyd and Cazalet [6], and Morris [21].

dissertation is illustrated in Figure 2.5. In this hypothetical organization there are three types of positions.

At the top of Figure 2.5 we have identified an impresario who is in some sense "running the show." More precisely, he is responsible for defining the structure of the organization and for defining the responsibilities of the other members of the organization.

On the left side of Figure 2.5 we have identified several entrepreneurs. Generally, the entrepreneurs are directly responsible for employing resources in the productive activities that the organization is organized to perform.

Finally, a resource manager is shown on the right side of Figure 2.5. The resource manager is responsible for satisfying the entrepreneurs requests for scarce resources or disposing of abundant resources. In more complicated situations there will be several resource managers; one for each resource.

This general organizational structure is applicable to a wide variety of situations. One example in a corporate context interprets the impresario as the company president, the entrepreneurs as division managers, and the financial vice-president as one of the resource managers. The same general structure can be applied at other levels in a corporation.

In a governmental organization, we might interpret the impresario as the governor of a state or the president of a country. The entrepreneurs might be lower-level decision makers or even independent citizens whose decisions interact through common resources. The resource manager might represent a public institution charged with the responsibility for the management of a scarce natural resource.

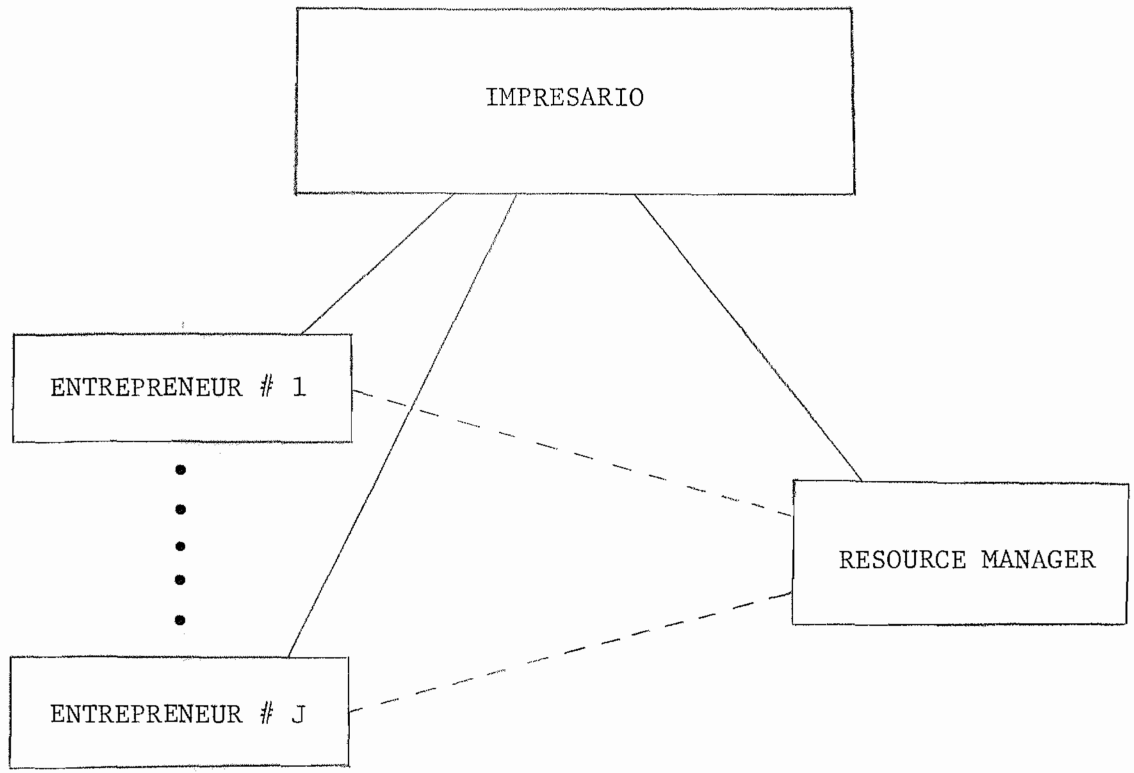


Figure 2.5: DECENTRALIZED ORGANIZATIONAL INTERPRETATION

We can interpret our single resource example in terms of the general decentralized organization that was illustrated in Figure 2.5. In this case, the entrepreneurs are project managers responsible for deciding on the amount of a resource to use in their project. Given the price of the resource, each project manager independently maximizes "profit" according to

$$\max_{x_j \in X_j} \left[r_j(x_j) - \lambda x_j \right]$$

where λx_j is the "penalty" paid for the use of the resource.

The resource manager's task depends on the algorithm chosen by the impresario or possibly by the resource manager himself.

For example, the successive approximations algorithm instructs the resource manager to set the price of the resource equal to the marginal cost of the resource.

With the price directive gradient algorithm, the resource manager acts more like a project manager. Independently of the project managers, he maximizes "profit" according to

$$\max_{y} \left[\lambda y - C(y) \right]$$

where λy is the "penalty" he receives from the entrepreneurs. The price of the resource on the next iteration depends on the "excess supply" of the resource. The excess supply is the difference between the resource manager's purchase of the resource in the resource market and the total requirements of the project managers. The price of the resource, in this case, could be set by either the impresario or the resource manager.

It is tempting to evaluate the performance of the project managers on the basis of their "profit,"

$$r_j(x_j^*) - \lambda x_j^* \ .$$

However, the distribution of the cost of the resource among the project managers is rather arbitrary. The arbitrariness of the distribution of resource cost will become more obvious in the discussion of penalty functions following later in this section. The appropriate basis on which to evaluate a manager's performance is the quality of his decisions rather than on the magnitude of his profits. In a highly interactive situation, a project manager's profit depends largely on factors beyond his control.

An Alternative Development of Decomposition

In this subsection we develop some of the results of the previous subsections under the assumption of differentiability. This alternative development is interesting because it relates the results of this dissertation to the optimization methods provided by calculus. However, in practical terms, the development is useful because it provides an approach to identifying resources and prices in very complex problems where unaided intuition is often misleading.

We recall that our simple resource allocation problem can be written as

$$\underset{\underline{x} \in X}{\text{maximize}} \left[\sum_{j=1}^{J} r_j(x_j) - C\left(\sum_{j=1}^{J} x_j \right) \right]$$

if we assume that the revenue function is separable. For simplicity, we will examine the problem without constraints, although some simple constraints such as positivity constraints $(\underline{x} \geq 0)$ could be treated easily. The necessary conditions[†] for $\underline{x}^*$ to be a maximum for this problem are

$$\frac{\partial}{\partial x_j} r_j(x_j)\bigg|_{x_j^*} - \frac{\partial}{\partial x} C(y)\bigg|_{y = \sum_{j=1}^{J} x_j^*} = 0 \quad \text{for} \quad j = 1, \ldots, J .$$

These conditions are also sufficient if the function

$$\sum_{j=1}^{J} r_j(x_j) - C\left(\sum_{j=1}^{J} x_j\right)$$

is concave. To determine the optimal allocation $\underline{x}^*$ we must solve the J simultaneous equations describing the necessary conditions.

The solution of the equations for $\underline{x}^*$ is made relatively easy if we observe that the term

$$\frac{\partial}{\partial y} C(y)\bigg|_{y = \sum_{j=1}^{J} x_j^*}$$

is independent of the index j. Suppose we set this term to an initial value, λ^0. Then the J simultaneous equations become J independent equations of the form

$$\frac{\partial}{\partial x_j} r_j(x_j)\bigg|_{x_j^n} - \lambda^n = 0$$

each of which can be solved for x_j^n. Having solved for $\underline{x}^n$ we can

[†] Background material for the development can be found in most elementary texts on calculus such as Courant [11].

calculate λ^{n+1} according to

$$\lambda^{n+1} = \frac{\partial}{\partial y} C(y) \bigg|_{y = \sum_{j=1}^{J} x_j^n} .$$

If $\lambda^{n+1} = \lambda^n$, then $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Obviously, the solution process just described is the successive approximations algorithm developed earlier in this section. In economic terms we are simply allocating resources so that marginal revenue equals marginal cost.

An important insight is that the allocation x_j^n satisfying the equation

$$\frac{\partial}{\partial x_j} r_j(x_j) \bigg|_{x_j^n} - \lambda^n = 0$$

is also satisfied by the solution to the problem,

$$\underset{x_j}{\text{maximize}} \left[r_j(x_j) - \lambda^n x_j \right]$$

under the assumptions made in this subsection. Thus, having defined the price λ, we can solve the original J-dimensional problem by solving J one-dimensional problems. Theorem I and the subsequent results based on that theorem demonstrate that this result is valid for a much larger class of problems than indicated by the assumptions in this subsection.

A complete development of decomposition under the assumption of differentiability is in Boyd and Cazalet [6].

A Method for Identifying Resources and Prices

In complex problems it is often difficult to identify the appropriate resources on which to define prices. In problems where there are strong interactions between projects, decomposition can often be achieved by defining new resources or parameters to characterize the interactions. The resulting decomposition is most effective if the resources are chosen to minimize the number of resource markets. One looks for resources that have an additive relationship among projects so that a price defined on a single resource serves to coordinate many projects.

A general approach towards identifying resources and prices is suggested by the methods of the previous subsection. The first step involves careful formulation of the problem as an unconstrained resource allocation problem. Next, we temporarily assume that all functions are differentiable and that they satisfy the appropriate concavity and convexity requirements. Then we differentiate the objective function with respect to each of the decision variables and set the result to zero. At this point we have a set of simultaneous equations describing the necessary conditions for an optimum.

To iteratively solve this set of equations we can guess certain terms in these equations and then solve for the optimal allocation. The selection of which terms to guess is a creative process whose success depends upon the skill of the analyst. Finally, we formalize the solution in the form of an algorithm. The results of this dissertation justify the application of the solution methods to problems where the required differentiability and convexity assumptions are not present.

Several examples of the application of this method of identifying resources and prices are contained in Boyd and Cazalet [6] and Boyd [5]. The decomposition of the electrical power system example in Chapter IV is a practical application of this method.

The general idea of defining new resources to account for interactions between projects is not original. In economics, the problem of interactions between decisions made in a market economy corresponds to interactions between our projects. Those interactions that are not appropriately priced by an economic market are called external or neighborhood effects. Although it is often overlooked, one way of accounting for these interactions is to define additional resources and then set prices on these resources.† Sometimes a new market mechanism can be developed to set the price, or the price can be set by political means. Usually, legislation is required to enforce payment for these new resources. In a centralized resource allocation problem many of the practical difficulties inherent in defining new resources for an economy, are absent.

Penalty Function Methods (Theorem I', Bounds, Algorithms and Discussion)

We have previously observed that the term λx_j in the project managers' problem

$$\underset{x_j \in X_j}{\text{maximize}}\ r_j(x_j) - \lambda x_j$$

can be viewed as the penalty paid for the use of the resource. In our

† Arrow [1] observes that externalities are simply a matter of the classification of resources.

discussion thus far, the penalty is a linear function of the amount of the resource employed. In this subsection we investigate general penalty functions where the penalty may be a nonlinear function of the amount of resource employed.

Nonlinear penalty functions are interesting for two reasons. First, the study of nonlinear penalty functions provides insight into linear penalty functions involving prices. Second, nonlinear penalty functions conceptually provide the means for resolving the problem of gaps. In practice, however, nonlinear penalty functions have limited value. The price of using nonlinear penalty functions is a reduction in the degree of decomposition.

The following theorem provides the theoretical foundation for general penalty function methods. Theorem I is a special case of this theorem and both theorems apply to the single resource problem.

THEOREM I': If $\underline{x}^*$ maximizes

$$R(\underline{x}) - P(\underline{x})$$

over all $\underline{x} \in X$, and if $\underline{y}^*$ maximizes

$$P(\underline{y}) - C \sum_{j=1}^{J} y_j \ ,$$

over all $\underline{y}$, and if

$$\underline{x}^* = \underline{y}^* \ ,$$

then $\underline{x}^*$ maximizes

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right)$$

over all $\underline{x} \in X$.

Proof: a) The theorem statement implies the following two inequalities:

$$R(\underline{x}) - P(\underline{x}) \leq R(\underline{x}^*) - P(\underline{x}^*) \tag{1}$$

holds for all $\underline{x} \in X$, and

$$P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right) \leq P(\underline{y}^*) - C\left(\sum_{j=1}^{J} y_j^*\right) \tag{2}$$

holds for all $\underline{y}$.

b) Combining inequalities (1) and (2) gives

$$R(\underline{x}) - P(\underline{x}) + P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right) \leq R(\underline{x}^*) - P(\underline{x}^*) + P(\underline{y}^*) - C\left(\sum_{j=1}^{J} y_j^*\right) \tag{3}$$

which holds for all $\underline{x} \in X$ and all $\underline{y}$.

c) Since (3) holds for all $\underline{y}$, it must also hold for $\underline{y} = \underline{x}$ where $\underline{x} \in X$. In this case, the terms involving $P(\ )$ on the left side of (3) cancel and

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right) \leq R(\underline{x}^*) - P(\underline{x}^*) + P(\underline{y}^*) - C\left(\sum_{j=1}^{J} y_j^*\right) \tag{4}$$

holds for all $\underline{x} \in X$.

d) By the statement of the theorem,

$$\underline{y}^* = \underline{x}^* .$$

Thus, the terms involving $P(\ )$ on the right side of (4) cancel and

$$R(\underline{x}) - C\left(\sum_{j=1}^{J} x_j\right) \leq R(\underline{x}^*) - C\left(\sum_{j=1}^{J} x_j^*\right) \tag{5}$$

holds for all $\underline{x} \in X$. Hence the theorem is proved.

This theorem is essentially Theorem I with $\lambda \sum_{j=1}^{J} x_j$ replaced by $P(\underline{x})$. The theorem provides only sufficient conditions for an optimal allocation and does not require any assumptions on the form of the functions other than real-valuedness.

Some insight into general penalty functions is provided by the graphical examples in Figure 2.6. As in Figure 2.3 we define

$$R(x) = \max_{\underline{x} \in X} R(\underline{x})$$
$$\text{subject to} \quad \sum_{j=1}^{J} x_j = x$$

so that a two-dimensional presentation can be used.

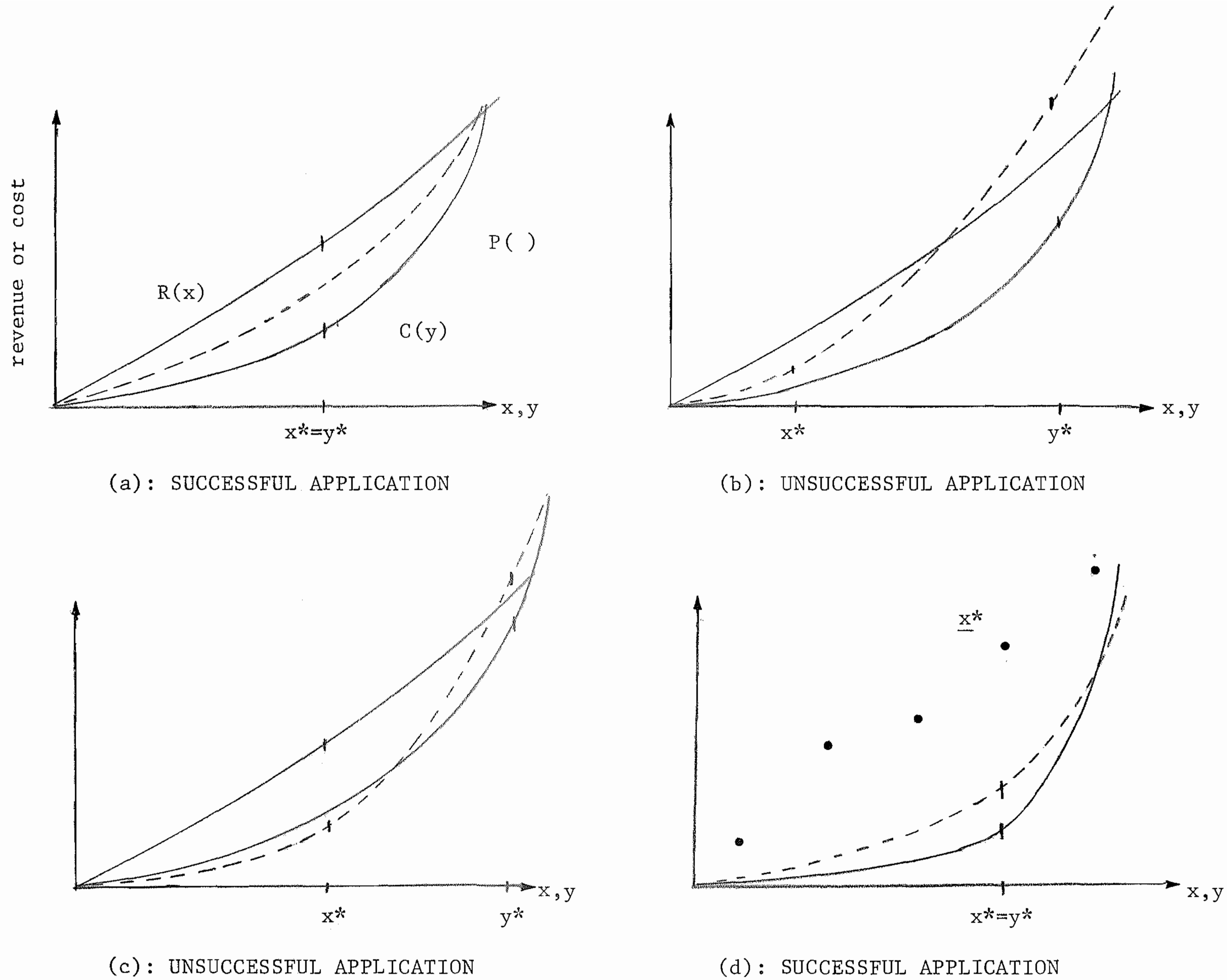
Consider Figure 2.6(a). A particular penalty function $P(y)$[†] is illustrated by the dotted line. The first maximization problem in Theorem I' is to maximize

$$R(x) - P(x)$$

where x must be chosen to satisfy the requirement that $\underline{x} \in X$. Graphically, the equivalent problem is to choose the allocation x that maximizes the vertical distance between $R(x)$ and $P(x)$. This maximum occurs at x^*. The second maximization problem in Theorem I' is to maximize

$$P(y) - C(y)$$

[†] This penalty function depends only on the total resources, y.



(a): SUCCESSFUL APPLICATION

(b): UNSUCCESSFUL APPLICATION

(c): UNSUCCESSFUL APPLICATION

(d): SUCCESSFUL APPLICATION

Figure 2.6: EXAMPLE OF THE APPLICATION OF THEOREM I'

over all y. Graphically, we maximize the vertical distance between $P(y)$ and $C(y)$ in Figure 2.6(a). The optimal allocation for this second subproblem is y^*. In this case $y^* = x^*$. Thus, all of the conditions of Theorem I' are satisfied and x^* must be the optimal allocation.

The problem illustrated in Figure 2.6(a) could not have been solved by linear penalty function methods. If $P(\ )$ were linear, then the dotted line would be straight. Under these conditions the solution to the first maximization problem cannot be the optimal solution to the overall problem. In a previous subsection we described this situation in terms of gaps. Figure 2.6(a) demonstrates that an appropriate choice of a penalty function can resolve the problem of gaps.

Figure 2.6(b) illustrates a case where the conditions of Theorem I' are not satisfied because x^* is not equal to y^*. In Figure 2.6(c) the application of the theorem is also unsuccessful. A successful application of the theorem to a discrete problem is illustrated in Figure 2.6(d).

General penalty functions also provide bounds on the optimal profit during the course of an iterative search. The lower bound is given by the best available solution obtained in the course of the iterations. The upper bound is given by inequality (4) in the proof of Theorem I. The bounds are stated formally below.

BOUNDS:

Let $\underline{x}'$ maximize

$$R(\underline{x}) - P(\underline{x})$$

over all $\underline{x} \in X$, and let $\underline{y}'$ maximize

$$P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right)$$

over all $\underline{y}$. Lower and upper bounds on the optimal profit are given by

$$p^{\ell} = R(\underline{x}') - C\left(\sum_{j=1}^{J} x_j'\right)$$

$$p^{u} = R(\underline{x}') - C\left(\sum_{j=1}^{J} y_j'\right) + P(\underline{x}') - P(\underline{y}') \ .$$

The basic algorithms for linear penalty functions carry over to nonlinear penalty functions with minor changes. The successive approximations algorithm is as follows:

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial penalty function $P^0(\ )$ subject to the conditions discussed below.
2. Maximize

$$R(\underline{x}) - P^n(\underline{x})$$

   over all $\underline{x} \in X$. Call the result $\underline{x}^n$.

3. Determine a new penalty function such that

$$\frac{\partial}{\partial y_j} P^{(n+1)}(\underline{y})\Big|_{\underline{y}=\underline{x}^n} = \frac{\partial}{\partial y} C(y)\Big|_{y = \sum_{j=1}^{J} x_j^n} \quad \text{for } j = 1, \ldots , J \ .$$

4. If $P^{n+1}(\underline{x}) = P^n(\underline{x})$ for all $\underline{x} \in X$, then, subject to the conditions discussed below, $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Otherwise return to Step 2 using $P(\ )^{n+1}$.

Linear penalty functions require that $C(\ )$ be convex and differentiable if the structure of the successive approximations algorithm is to satisfy conditions of Theorem I. The conditions of Theorem I' are satisfied for nonlinear penalty functions if the function

$$P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right)$$

is concave and differentiable. If this condition is met, then the J equations in Step 3 of the algorithm provide necessary and sufficient conditions for $\underline{y}^n$ to be the global solution to the following problem:

$$\underset{\underline{y}}{\text{maximize}}\ P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right) .$$

For linear penalty functions the conditions are equivalent to assuming convexity and differentiability for $C(\ )$.

The conditions on $P(\ )$ and $C(\ )$ point out one of the inherent practical difficulties associated with penalty functions. In order to use penalty functions to probe gaps in a revenue function, the function $P(\ )$ must be convex. If the combined function

$$P(\underline{y}) - C\left(\sum_{j=1}^{J} y_j\right)$$

is to be concave, then, loosely speaking $C(\ )$ must be "more convex" than P( ). It is not always easy to find penalty functions with the "right degree of convexity."

An alternative algorithm that satisfies the conditions of Theorem I' without restrictive assumptions is the penalty directive gradient algorithm. This algorithm can only be described in terms of a parameterized set of penalty functions. Let $P(\ |\underline{\beta})$ be a set of penalty functions with parameters $\underline{\beta} = (\beta_1, \ldots, \beta_N)$. The penalty directive gradient algorithm is as follows:

PENALTY DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial set of parameters $\underline{\beta}^0$.

2. Maximize

$$R(\underline{x}) - P(\underline{x}|\underline{\beta}^n)$$

   over all $\underline{x} \in X$. Call the result $\underline{x}^n$.

3. Maximize

$$P(\underline{y}|\underline{\beta}^n) - C\left(\sum_{j=1}^{J} y_j\right)$$

   over all $\underline{y}$. Call the result $\underline{y}^n$.

4. If $\underline{x}^n = \underline{y}^n$, then the conditions of Theorem I' are satisfied and $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal solution. Otherwise, compute a new set of parameters according to

$$\beta_m^{n+1} = \beta_m^n + \alpha \frac{\partial}{\partial\beta_m} P(\underline{x}|\underline{\beta})\bigg|_{\underline{\beta}^n} - \frac{\partial}{\partial\beta_m} P(\underline{y}|\underline{\beta})\bigg|_{\underline{\beta}^n} \quad .$$

This algorithm can be derived by applying a gradient search to the problem of minimizing the upper bound. The upper bound is given by

$$p^u = R(\underline{x}') - C\left(\sum_{j=1}^{J} y_j'\right) + P(\underline{x}'|\underline{\beta}) - P(\underline{y}'|\underline{\beta}) .$$

The rate of change (gradient) of the upper bound with respect to the parameter β_m is given by

$$\frac{\partial p^u}{\partial \beta_m} = \frac{\partial}{\partial \beta_m} P(\underline{x}'|\underline{\beta}) - \frac{\partial}{\partial \beta_m} P(\underline{y}'|\underline{\beta}) .$$

The parameter adjustment formula in Step 4 of the algorithm moves the vector of parameters in the direction of the gradient. In the case of linear penalty functions only a single parameter, the price λ, is required.

The practical application of penalty function methods involves compromises. From the point of view of decomposition, nonlinear penalty functions are less desirable than linear penalty functions. A degree of decomposition can be obtained by using separable penalty functions. For example,

$$P(\underline{x}) = \sum_{j=1}^{J} p_j(x_j)$$

is separable over the set of projects. The penalty function $p_j(x_j)$ involves only the amount of resource consumed by the j^{th} project. Thus, given the penalty functions, the allocation decisions can be made independently. Unfortunately, the informational and computational costs associated with N nonlinear penalty functions are much greater than the costs associated with N linear penalty functions where only a single parameter is involved.

The practical difficulties associated with penalty functions limit their usefulness. In problems where gaps are important, penalty functions may provide a solution. Sometimes restructuring the problem or looking at the problem from a different point of view is superior to implementing nonlinear penalty functions.

The subject of penalty functions is also considered in the literature on constrained optimization. The general idea is to replace constraints with penalty functions and then adjust the parameters of the penalty function until the original constraints are satisfied. The Lagrangian method described earlier in this section is recognized as a special case of penalty function methods. Fiacco and McCormick [14] provide a comprehensive treatment of penalty function methods in constrained problems. Bellmore, Greenberg and Jarvis [3] provide a clear discussion of penalty functions that includes a discussion of gaps. Arrow and Hurwicz [2] discuss penalty functions from the point of view of economic theory.

In constrained problems, penalty functions are useful because they convert constrained problems into a series of unconstrained optimization problems. Often the unconstrained problem is easier to solve. Generally, decomposition is feasible only for linear penalty functions. In the class of problems treated in this dissertation, the problem is formulated as an unconstrained problem. Any computational benefits from penalty functions applied to unconstrained problems must result from decomposition.

2.2 Multiple Resource Problems with Separable Objective Functions

Generally, the results for single resource problems carry over directly to multiple resource problems. In this section we will restate the major results of Section 2.1 in terms of a multi-resource example.

The Example

Consider the problem of allocating amounts of K resources among J projects. Let

x_{jk} ≡ amount of the k^{th} resource used by the j^{th} project where $j = 1, \ldots, J$ and $k = 1, \ldots, K$.

$\underline{x}_k \equiv (x_{1k}, \ldots, x_{Jk})$, a vector.

$\underline{x}_j \equiv (x_{j1}, \ldots, x_{jK})$, a vector.†

$$\underline{x} \equiv \begin{bmatrix} x_{11}, x_{12}, \ldots, x_{1K} \\ x_{21}, x_{22}, \ldots, x_{2K} \\ \vdots \\ x_{J1}, x_{J2}, \ldots, x_{JK} \end{bmatrix}, \text{ a matrix.}$$

$R(\underline{x})$ ≡ total revenue from all projects as a function of the amount of each resource employed by each project. In general,

$$R(\underline{x}) = \sum_{j=1}^{J} r_j(\underline{x}_j) .$$

y_k ≡ total amount of the k^{th} resource used by all J projects. Thus

$$y_k = \sum_{j=1}^{J} x_{jk}, \quad k = 1, \ldots, K .$$

$\underline{y} \equiv (y_1, \ldots, y_K)$, a vector.

$C(\underline{y})$ ≡ total cost of all resources purchased in the resource market.

† The distinction between $\underline{x}_j$ and $\underline{x}_k$ will usually be implied by the context in which they are used in an equation.

The resource allocation problem is to choose an $\underline{x}$ to maximize

$$R(\underline{x}) - C(\underline{y})$$

where $\underline{x}$ is chosen from a completely arbitrary set X.

Mathematical Results (Theorem II, Bounds, and Algorithms)

THEOREM II: If $\underline{x}^*$ maximizes

$$R(\underline{x}) - \sum_{k=1}^{K} \lambda_k \left[\sum_{j=1}^{J} x_{jk} \right]$$

over all $\underline{x} \in X$, and if $\underline{y}_k^* = \sum_{j=1}^{J} x_{jk}$, $k = 1, \ldots, K$ maximizes

$$\sum_{k=1}^{K} \lambda_k y_k - C(\underline{y})$$

over all $\underline{y}$, then $\underline{x}^*$ maximizes

$$R(\underline{x}) - C(\underline{y})$$

over all $\underline{x} \in X$.

The proof of this theorem is a straightforward extension of Theorem I.

BOUNDS:

Let $\underline{x}'$ maximize

$$R(\underline{x}) - \sum_{k=1}^{K} \lambda_k \left[\sum_{j=1}^{J} x_{jk} \right]$$

over all $\underline{x} \in X$ and let $\underline{y}'$ maximize

$$\sum_{k=1}^{K} \lambda_k y_k - C(\underline{y})$$

over all $\underline{y}$. A lower and upper bound on the optimal profit are given by

$$p^{\ell} = R(\underline{x}') - C\left(\sum_{j=1}^{J} x'_{j1}, \sum_{j=1}^{J} x'_{j2}, \ldots, \sum_{j=1}^{J} x'_{jk}\right),$$

and

$$p^{u} = R(\underline{x}') - C(\underline{y}') + \sum_{k=1}^{K} \lambda_k \left[y'_k - \sum_{j=1}^{J} x'_{jk} \right].$$

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial price vector, $\underline{\lambda}^0 = (\lambda_1, \ldots, \lambda_K)$ or start with a trial $\underline{y}$ in Step 3.
2. Maximize

$$R(\underline{x}) - \sum_{k=1}^{K} \lambda_k \left[\sum_{j=1}^{J} x_{jk} \right]$$

over all $\underline{x} \in X$. Call the result $\underline{x}^n$.

3. Calculate $\underline{\lambda}^{n+1}$ according to

$$\lambda_k = \frac{\partial}{\partial y_k} C(\underline{y}) \Bigg|_{\substack{y_k = \sum_{j=1}^{J} x'_{jk} \\ k=1,\ldots,K}}, \quad k = 1, \ldots, K.$$

4. If $\underline{\lambda}^{n+1} = \underline{\lambda}^n$, then the conditions of Theorem II are satisfied and $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Otherwise, return to Step 2 using $\underline{\lambda}^{n+1}$.

(Note: $C(\underline{y})$ must be convex and differentiable for this algorithm.)

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial price vector, $\lambda^0 = (\lambda_1, \ldots, \lambda_K)$.

2. Maximize

$$R(\underline{x}) - \sum_{k=1}^{K} \lambda_k \left[\sum_{j=1}^{J} x_{jk} \right]$$

over all $\underline{x} \in X$. Call the result $\underline{x}^n$.

3. Maximize

$$\sum_{k=1}^{K} \lambda_k y_k - C(\underline{y})$$

over all $\underline{y}$. Call the result $\underline{y}^n$.

4. If $\sum_{j=1}^{J} x_{jk}^n = y_k^n$ for $k = 1, \ldots, K$, then the conditions of Theorem II are satisfied and $\underline{x}^n$ is equal to $\underline{x}^*$, the optimal allocation. Otherwise, compute $\underline{\lambda}^{n+1}$ according to

$$\lambda_k^{n+1} = \lambda_k^n - \alpha \left[y_k^n - \sum_{j=1}^{J} x_{jk}^n \right] \qquad k = 1, \ldots, K$$

(where α is an appropriately chosen constant) and return to Step 2.

Decomposition

In Section 2.1, decomposition was possible when the projects and the set X had certain special structure. For multiple resources, decomposition of the project decisions is possible under the same conditions. The required conditions are that the projects be independent,

$$R(\underline{x}) = \sum_{j=1}^{J} r_j(\underline{x}_j)$$

and

$$X = X_1 \times X_2 \times \cdots \times X_{J-1} \times X_J$$

where

$$\underline{x}_j \in X_j \quad \text{and} \quad \underline{x}_i \notin X_j \quad \text{for} \quad i \neq j \ .$$

Under these conditions, Step 2 of the algorithms becomes

$$\max_{\underline{x} \in X} \left[R(\underline{x}) - \sum_{k=1}^{K} \lambda_k \left[\sum_{j=1}^{J} x_{jk} \right] \right] = \sum_{j=1}^{J} \max_{\underline{x}_j \in X_j} \left[r_j(\underline{x}_j) - \sum_{k=1}^{K} \lambda_k x_{jk} \right] .$$

In the general case, the resource market cost functions are dependent and decomposition of one resource market from another is not directly possible. When decomposition of the resource markets is not possible, then, for example, Step 3 of the price directive gradient algorithm involves a multi-dimensional search. Decomposition of the resource markets is possible when

$$C(\underline{y}) = \sum_{k=1}^{K} c_k(y_k)$$

in which case Step 3 of the price directive gradient algorithm involves only one-dimensional searches. Sometimes $C(\underline{y})$ can be partially decomposed. If a complete decomposition is not directly possible, then it is often worthwhile to restructure the problem to permit further decomposition.

Decision Variables

In the present formulation of our examples we have focused on the resource allocation problem. In many problems, however, it is useful to focus on the decisions that control the allocation of resources. In this subsection we will develop some notation that emphasizes the distinction between a decision and the eventual allocation of resources that depend on that decision. This distinction is particularly important in problems involving time or uncertainty.

Let

$\underline{\theta}_j$ $\equiv$ vector of decision variables (policy) associated with the j^{th} project. The number of elements in $\underline{\theta}_j$ need not be defined at this point.

$x_{jk}(\underline{\theta}_j)$ $\equiv$ amount of the k^{th} resource used by the j^{th} project as a function of $\underline{\theta}_j$.

$\underline{\theta}$ $\equiv (\underline{\theta}_1, \ldots, \underline{\theta}_j)$, a matrix giving the decision policy for the problem.

Θ $\equiv$ set of all possible policies.

Using decision variables the resource allocation problem treated in this section becomes

$$\max_{\underline{\theta} \in \Theta} \left[R(\underline{\theta}) - C(\underline{y}(\underline{\theta}))\right]$$

$$\text{where} \quad y_k(\underline{\theta}) = \sum_{j=1}^{J} x_{jk}(\underline{\theta}_j) \qquad k = 1, \ldots, K .$$

The total project revenue in the above formulation is now a function of the project decision variables. Decomposition among projects is possible when

$$R(\underline{\theta}) = \sum_{j=1}^{J} r_j(\underline{\theta}_j)$$

and

$$\Theta = \Theta_1 \times \cdots \times \Theta_J .$$

2.3 Problems with Arbitrary Objective Functions

In this section we develop methods for treating resource allocation problems where the objective function does not easily separate into revenue and cost terms. In practical situations, arbitrary objective functions are often the result of multiple measures of performance.† For example, monetary profit may not be the only consideration in evaluating a resource allocation. If the decision maker's value function is defined over a number of measures of performance or outcome variables, then the methods developed in this section are useful.

The results of this section show that problems with arbitrary objective functions can be treated using methods similar to those developed for problems with separable objective functions in Sections 2.1 and 2.2. The optimality theorem developed in this section provides the theoretical basis for transforming problems with arbitrary objective functions into problems with separable objective functions.

A fundamental approach to problems with complex preferences is in Boyd [5]. He develops a methodology for assessing preferences in complicated situations. In this dissertation we simply intend to demonstrate that problems with arbitrary objective functions are conceptually no more difficult than problems with separable objective functions.

---

† We use the term "arbitrary" rather than the term "nonseparable" to describe the objective function here because the methods of this section also apply to and provide insight to problems with separable objective functions as a trivial case.

Introduction to Ordinal Value Functions

An ordinal value function is used to encode a decision maker's attitude towards the outcome of a deterministic resource allocation problem. An arbitrary objective function usually is the direct result of an ordinal value function or can be interpreted in terms of an ordinal value function. This subsection provides a nonrigorous introduction to ordinal value functions as motivation for the theoretical study of decomposition under arbitrary objective functions.

The preferences of a decision maker can be described in terms of the resources that can be identified in a problem. Only some of the resources in a complex problem are of direct concern to the decision maker. Resources can be classified as either primary or secondary resources. Primary resources are those resources directly consumed or valued by the decision maker. Secondary resources are resources that are indirectly valued because they are useful in producing primary resources.

Ordinal value functions are defined on the primary resources of a problem. A useful graphical device in the study of ordinal value functions is the indifference curve. Indifference curves are isovalue curves defined on the space of primary resources.

Figure 2.7 provides an elementary example of indifference curves. In this example the decision maker desires more of both resources x_1 and x_2. Any two resource allocations lying on the same indifference curve are said to have "equivalent value" to the decision maker.

Conceptually, indifference curves can be encoded by questioning the decision maker. Only questions involving comparisons between pairs

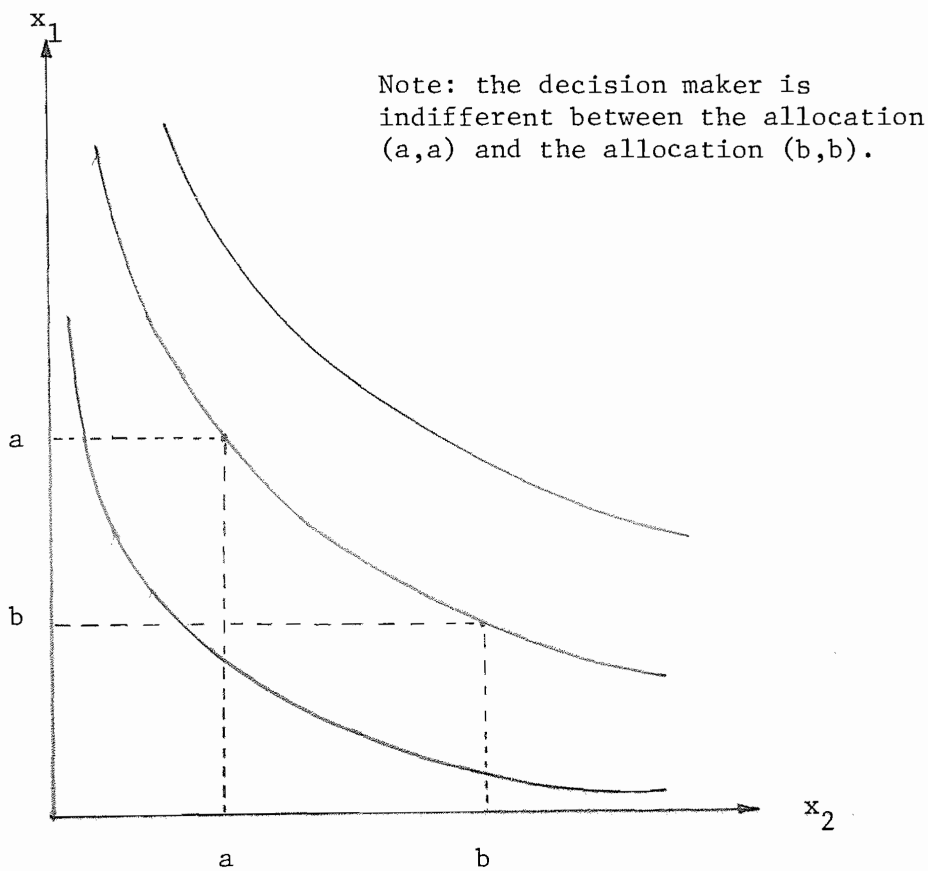


Figure 2.7: INDIFFERENCE CURVES

of resources are necessary. The preferences of the decision maker can be completely encoded without defining an index or cardinal value function which would assign unique numerical values to the resource allocations.

Indifference curves specify a ranking of alternative resource allocations. This ordering of the resource allocations is said to be an ordinal ranking because there is no need to assign a unique numerical value to the resource allocations.

From a computational point of view, it is often easier to maximize an objective function than it is to work directly with indifference curves. As long as the ordering of the ranking provided by the indifference curves in unchanged we can arbitrarily assign numerical values to the curves. If a function is used to assign the numerical values it is called an ordinal value function. Usually, we can arrange things so that the most preferred resource allocation also maximizes the ordinal value function.

In problems where both monetary and other resources are involved, the ordinal value function can be expressed in monetary terms. Where money is not involved, the ordinal value function can be expressed in equivalent units of one of the resources. In problems where time is involved the ordinal value function can be expressed in terms of an equivalent uniform flow.

This approach to the encoding of complex preferences has evolved out of the economist's use of these mathematical tools. For a more rigorous discussion of indifference curves and ordinal value functions from the general point of view of this introduction see Raiffa [25], Pollard [24], Boyd and Matheson [8], and Boyd [5].

In this section the notation

$$V(\underline{z})$$

is used to denote an ordinal value function defined on the resource vector z.† The variables z and w denote primary resources while the variables x and y denote secondary resources.

The Example

Consider the problem of selecting a resource allocation $\underline{z}$ to maximize an ordinal value function $V(\underline{z})$. We will employ the decision variable notation introduced in Section 2.2. Let

$\underline{\theta}$ ≡ vector of decision variables (policy).

Θ ≡ set of all possible policies.

$z_k(\underline{\theta})$ ≡ amount of the k^{th} primary resource as a function of the policy $\underline{\theta}$. The problem structure relating decision variables to secondary resources and secondary resources to primary resources is imbedded in this function.

$\underline{z}(\underline{\theta})$ ≡ vector of primary resources as a function of the policy $\underline{\theta}$. Thus,

$$\underline{z}(\underline{\theta}) = (\underline{z}_1(\underline{\theta}), \ldots , \underline{z}_k(\underline{\theta})) .$$

The resource allocation problem is to choose a policy $\underline{\theta} \in \Theta$ to

† The indifference curves corresponding to $V(\underline{z})$ are given by values of $\underline{z}$ satisfying the equation, $V(\underline{z})$ = constant, where the constant is the numerical value assigned to an indifference curve.

maximize

$$V(\underline{z}(\underline{\theta})) .$$

The simplicity of the problem statement hides the structure imbedded in this problem. Within this structure, secondary resources can be identified. Later, we will consider examples where the function $\underline{z}(\underline{\theta})$ is separable into project revenue and resource market cost functions. Most of the results of this section can be developed without assuming a special structure for $\underline{z}(\underline{\theta})$.

Mathematical Results (Theorem III, Bounds, and Algorithms)

The following theorem applies to the example discussed in this section:

THEOREM III: If $\underline{\theta}^*$ maximizes

$$\sum_{k=1}^{K} \mu_k z_k(\underline{\theta})$$

over all $\theta \in \Theta$, and if $\underline{w}^*$ maximizes

$$V(\underline{w}) - \sum_{k=1}^{K} \mu_k w_k$$

over all $\underline{w}$, and if

$$\underline{w}^* = \underline{z}(\underline{\theta}^*)$$

then $\underline{\theta}^*$ maximizes

$$V(\underline{z}(\underline{\theta}))$$

over all $\underline{\theta} \in \Theta$.

Proof: a) The theorem statement implies the following two inequalities:

$$\sum_{k=1}^{K} \mu_k z_k(\underline{\theta}) \leq \sum_{k=1}^{K} \mu_k z_k(\underline{\theta}) \tag{1}$$

holds for all $\underline{\theta} \in \Theta$, and

$$V(\underline{w}) - \sum_{k=1}^{K} \mu_k w_k \leq V(\underline{w}^*) - \sum_{k=1}^{K} \mu_k w_k^* \tag{2}$$

holds for all $\underline{w}$.

b) Combining inequalities (1) and (2) gives

$$V(\underline{w}) - \sum_{k=1}^{K} \mu_k [w_k - z_k(\underline{\theta})] \leq V(\underline{w}^*) - \sum_{k=1}^{K} \mu_k [w_k^* - z_k(\underline{\theta}^*)] \tag{3}$$

which holds for all $\underline{w}$ and all $\underline{\theta} \in \Theta$.

c) Since (3) holds for all $\underline{w}$, it must also hold for $\underline{w} = \underline{z}(\underline{\theta})$ where $\underline{\theta} \in \Theta$. In this case the terms involving μ_k on the left side of (3) cancel and

$$V(\underline{z}(\underline{\theta})) \leq V(\underline{z}(\underline{\theta}^*)) - \sum_{k=1}^{K} \mu_k [w_k^* - z_k(\underline{\theta}^*)] \tag{4}$$

holds for all $\underline{\theta} \in \Theta$.

d) By the statement of the theorem

$$\underline{w}^* = \underline{z}(\underline{\theta}^*) \ .$$

Thus, terms involving μ_k on the right side of (3) cancel and the inequality

$$V(\underline{z}(\underline{\theta})) \leq V(\underline{z}(\underline{\theta}))$$

holds for all $\underline{\theta} \in \Theta$. Hence, the theorem is proved.

This theorem is a special case of Theorem II with minor changes in notation and with $R(\underline{x}) \equiv 0$ and $C(\underline{y}) \equiv -V(\underline{w})$. Thus, the discussion of Theorems I and II also applys to Theorem III.

Since Theorem III is not meaningful for a single resource, its geometric interpretation is more difficult to visualize. In the case where $k = 2$ the objective function $V(\ )$ defines a surface; the height of the surface above the (z_1, z_2) plane is given by the numerical value of the function. The set of resource allocations are points in the (z_1, z_2) plane. The prices μ_1 and μ_2 define the slope of a plane in three-dimensional space.

The first maximization problem in Theorem III is equivalent to raising the plane from below the (z_1, z_2) plane until it touches an element of the set of resource allocations in the (z_1, z_2) plane. The second maximization problem is equivalent to lowering the same plane until it touches the surface $V(\ )$. If the resource allocations determined by both maximization problems are the same, then the conditions of the theorem are satisfied. Naturally, gaps may exist that prevent the theorem from identifying the optimal solution.

The upper and lower bounds on the optimal value of the ordinal objective function can be obtained during an iterative search. The lower bound is given by the best available solution obtained in the course of the iterations. The upper bound is given by inequality (4) in the proof of Theorem III. The bounds are stated formally below:

BOUNDS:

Let $\underline{\theta}'$ maximize

$$\sum_{k=1}^{K} \mu_k z_k(\underline{\theta})$$

over all $\underline{\theta} \in \Theta$ and let $\underline{w}'$ maximize

$$V(\underline{w}) - \sum_{k=1}^{K} \mu_k w_k$$

over all $\underline{w}$. Lower and upper bound on the optimal value of $V(\ )$ are given by

$$V^{\ell} = V(\underline{z}(\underline{\theta}'))$$

$$V^{u} = V(\underline{w}') - \sum_{k=1}^{K} \mu_k (w_k' - z_k(\underline{\theta}')) \ .$$

The successive approximations algorithm is particularly applicable to problems with arbitrary objective functions because an ordinal value function is usually both differentiable and concave. The successive approximations algorithm follows from the statement of the necessary and sufficient conditions for a solution to the second optimization problem in Theorem III. The algorithm is as follows:

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial $\underline{\mu}$, $\underline{\mu}^0$ or start with a trial $\underline{z}$ at Step 3.
2. Maximize

$$\sum_{k=1}^{K} \mu_k^n z_k(\underline{\theta})$$

   over all $\underline{\theta} \in \Theta$. Call the result $\underline{\theta}^n$.

3. Calculate a new vector $\underline{\lambda}^{n+1}$ according to the relationship

$$\mu_k^{n+1} = \frac{\partial}{\partial w_k} V(\underline{w}) \bigg|_{\underline{w}=\underline{z}(\underline{\theta}^n)} \qquad k = 1, \ldots , K .$$

(Note: $V(\ )$ must be concave and differentiable in this algorithm.)

4. If $\underline{\mu}^{n+1} = \underline{\mu}^n$, then the conditions of Theorem III are satisfied and $\underline{\theta}^n$ is equal to $\underline{\theta}^*$, the optimal policy. Otherwise, return to Step 2 using $\underline{\mu}^{n+1}$.

A relaxation constant can be applied in Step 3 to improve convergence. In this case the new $\underline{\mu}$ is determined by the relationship

$$\mu_k^{n+1} = \alpha \frac{\partial}{\partial w_k} V(\underline{w}) \bigg|_{\underline{w}=\underline{z}(\underline{\theta}^n)} + (1-\alpha)\mu_k^n .$$

The price directive gradient algorithm is derived by applying a gradient algorithm to minimize the upper bound on the arbitrary objective function. The algorithm is as follows:

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial $\underline{\mu}$, $\underline{\mu}^0$.
2. Maximize

$$\sum_{k=1}^{K} \mu_k^n z_k(\underline{\theta})$$

over all $\theta \in \Theta$. Call the result $\underline{\theta}^n$.

3. Maximize

$$V(\underline{w}) - \sum_{k=1}^{K} \mu_k w_k$$

over all $\underline{w}$, Call the result $\underline{w}^n$.

4. If $\underline{w}^n = \underline{z}(\underline{\theta}^n)$, then the conditions of Theorem III are

satisfied and $\underline{\theta}^n$ is equal to $\underline{\theta}^*$, the optimal allocation. Otherwise, compute a new value of $\underline{\mu}$ according to

$$\mu_k^{n+1} = \mu_k^n + \alpha[w_k^n - z_k(\underline{\theta}^n)]$$

where α is an appropriate constant and return to Step 2.

Decomposition

The computational effectiveness of the algorithms depends on the relative difficulty of the original optimization problems and the optimization problems imbedded in the algorithms. The optimization problem in Step 2 of the algorithms has a special structure that the original problem does not possess; the objective function of the problem in Step 2 is separable. For problems with separable objective functions we can draw on the theory developed in Sections 2.1 and 2.2.

To see how the theory for separable objective function can be applied at this point, consider the case where

$$z_k(\underline{\theta}) = \sum_{j=1}^{J} r_{jk}(\underline{\theta}_j) - c_k\left(\sum_{j=1}^{J} x_j(\underline{\theta}_j)\right),$$

for $k = 1, \ldots, K$. The functions $r_{jk}(\underline{\theta}_j)$ describe the amount of the primary resource z_k produced by project j. The projects consume secondary resources x_j purchased in a resource market. The cost of the secondary resources in terms of the primary resource is given by $c_k(\ )$. In order to maximize

$$\sum_{j=1}^{J} r_{jk}(x_j(\underline{\theta}_j)) - c_k\left(\sum_{j=1}^{J} x_j(\underline{\theta}_j)\right)$$

over all $\underline{\theta} \in \Theta$, we can instead maximize

$$r_{jk}(\underline{\theta}_j) - \lambda_k x_j(\underline{\theta}_j)$$

over all $\underline{\theta} \in \Theta$ where $\quad = \Theta_1 \times \cdots \times \Theta_J$. The price λ_k can be determined along with the price μ_k in Step 3 of the successive approximations algorithm or Step 4 of the price directive gradient algorithm. Another alternative is to determine the prices on the primary resources by the successive approximations algorithm and use the price directive gradient algorithm to determine the prices on the secondary resources.

Organizational Interpretation

The results of this section can be interpreted in terms of a decentralized organization similar to that discussed in Section 2.1.

The role of the impresario who is at the head of the organization can be compared to the role of the resource managers. In Section 2.1, the impresario was responsible for the organization, but delegated the decision making to project managers and resource managers. With arbitrary objective functions, the impresario assumes a more active role similar to that of a resource manager.

In terms of the successive approximations algorithm the impresario's task is to assign prices on the primary resources. If the impresario has a complete description of the arbitrary objective function (ordinal value function) then he sets the price of the resource equal to the marginal value of the resource. If the impresario does not have a complete description of the arbitrary objective function then it may be easier for him to directly assign the prices rather than encode an ordinal objective function.[†] Thus, the impresario can be viewed as a primary resource manager. Of course, he could also delegate this task to a primary resource manager.

In a corporation the primary resources might be dividends, for example. The flow of dividends over time would depend on the flow of secondary resources over time. In this case, the arbitrary objective function (ordinal value function) would describe the corporation's time preference for dividends. The resulting prices can be given an interpretation as discount factors.

In a governmental example the primary resources would include measures of social value. The social values would depend on the allocation of other secondary resources. In this case, iteratively encoding the prices might be simpler.

2.4 Relationship of the Mathematical Foundation to the Methodology

This chapter provides a mathematical foundation for the methodology except for problems under uncertainty which are considered in Chapter IV. We will now discuss how the mathematical results fit into the overall methodology.

Using the methodology a decision problem is analyzed in two main steps. The first step is to carefully structure the problem and to identify the special structure required for decomposition. The second step is to justify the decomposition methods resulting from the first

---

[†] The observation that preferences can be encoded directly in terms of prices provides an approach to the assessment of preferences in complex situations. Often it is easier to iteratively assess the prices rather than to directly encode an ordinal value function. This approach is developed in detail by Boyd [5].

step by applying the mathematical foundations developed in this chapter.

The first step of the methodology requires creativity on the part of the analyst. Nevertheless, a general approach to structuring a problem is roughly as follows: First, formulate the decision problem as an unconstrained optimization problem. This means that the sources of resources are carefully modeled rather than simply limiting the availability of the resource. Also, the preferences of the decision maker are carefully structured rather than eliminating certain outcomes from consideration. Then, determine the necessary conditions for an optimal solution to the problem by temporarily assuming the methods of elementary calculus are valid for the problem. Generally, this temporary assumption will not be valid, but the mathematical results developed in this chapter can be applied later to justify the resulting decomposition.

As indicated in Section 2.1, the necessary conditions for an optimal solution to a decision problem result in a set of simultaneous equations. These equations usually can be solved iteratively by guessing certain terms in the equations and solving the equations one-by-one. The terms initially guessed are calculated from the solutions to the equations. If the guesses are correct then the optimal solution is available; otherwise use the new values of these terms to solve the equations again. Generally, the terms which are initially guessed can be interpreted as prices.

The iterative solution of the necessary conditions suggests a successive approximations algorithm for solving the problem. At this point, the analyst may review his formulation of the problem and method of solution. Often, by viewing the problem in another way or by solving the simultaneous equations differently, a better successive approximations algorithm can be devised. Generally, the analyst should attempt to minimize the number of resource markets in the problem so that only a few prices are used to coordinate the solution of a large number of subproblems or projects. The organizational interpretations of decomposition are particularly useful in providing insight into new ways to structure and solve a problem.

The second step in the methodology justifys the solution method in situations where the assumptions used in obtaining the necessary conditions are not valid. The mathematical results developed in this chapter can be applied by combining the results of the various theorems to solve the more complicated problems. In addition to justifying a method of solving a given problem, the mathematical results also provide additional algorithms and bounds on the value of the optimal solution.

Another way the mathematical results can be applied is to derive a new optimality theorem for the particular problem being analyzed. Usually, an optimality theorem can be stated once the successive approximations algorithm is developed in the first step of the methodology. The proofs of the optimality theorems developed in this chapter are simple and new optimality theorems can be easily proved for particular problems by following the same procedure used in proving the present theorems.

We have seen that the methodology developed in this dissertation combines the creativity of the analyst with a mathematical foundation for decomposition. In the next chapter, we demonstrate this methodology on a complex electrical power system problem.

CHAPTER III

AN ELECTRICAL POWER SYSTEM PLANNING PROBLEM

A dual purpose is intended for the example presented in this chapter. The first purpose of the example is to provide an effective medium for communicating some aspects of the methodology developed in this dissertation. The second purpose of the example is to develop new power system planning methods.

This chapter is written so that a person unfamiliar with the technical details of power systems can follow the logical development of the model. In Section 3.1 we point out that power system planning is complicated by the technical interactions between plants. These interactions are visible as the transmission lines which interconnect generating plants to serve a common market for electricity. In Section 3.2, the previous work on the example is discussed. Section 3.3 develops the model of the electrical system in detail. The interactions between plants are highlighted in the development of the model. The structure of the model permits very general submodels of the generating plants and other elements of the system.

In Section 3.4 decomposition of the problem is discussed. An important part of this section concerns the development of a sequential algorithm to overcome the effect of gaps without requiring penalty functions.

A numerical example is presented in Section 3.5. The data assumptions and computer program are explained. The results of the numerical example clearly demonstrate the practical value of the methodology developed in this dissertation. Finally, Sections 3.6 and 3.7 summarize the conclusions based on the model and suggest several directions for extending the scope of the model.

3.1 Introduction to Electrical Power System Planning

Electrical power system planning includes both strategic investment decisions and tactical operating decisions. In the analysis of investment decisions it is only necessary to treat the strategic decisions in detail. Tactical decisions need be considered only to the extent that they affect strategic decisions. In this chapter we focus on strategic decisions concerned with the installation of new generating plants.

Power system planning is a complex problem. The complexity of the problem is due, in part, to the following considerations:

1. Extended planning horizons are required in an analysis because of the long lifetimes of the expensive capital equipment used in power systems.
2. A detailed model of the entire system is required to evaluate capacity investments because of the complex technical interactions between generating plants in an interconnected power system.

For example, consider capital investments in generating equipment. The decision involves selection of a mix of plant types and plant sizes to be installed in an existing system. An analysis of such decisions requires extended planning horizons because some types of generating equipment operate for sixty years or longer. Technical interactions occur, for instance, because the quantity of fuel burned by a plant depends on the operation of all other plants by virtue of their interconnection to serve a common market for electricity. Technical interactions also occur because the system reliability depends in a complex way on the reliability of each plant.

The large number of possible combinations of individual plants in the long planning period makes optimization very difficult in power system planning. Decomposition and iteration are particularly useful in resolving such combinatorial problems. To achieve decomposition requires careful structuring of the problem to account for the interactions between plants. Before structuring the power system model we will summarize the previous work on the example.

3.2 Introduction to the Planning Example

Development of a detailed planning model for an electrical power system requires many man-years of effort. For this reason, the planning example is based on a model developed for a previous analysis of the same problem. The contribution of this chapter is to reformulate the model so that it can be decomposed.

The original analysis was performed for the government of Mexico [10]. The analysis considered the capacity expansion of the Mexican electrical power system. The emphasis of the analysis was on the development of a flexible planning tool and the installation of nuclear power plants in the middle 1970's.

The original analysis involved representatives of the Mexican electrical system and decision analysts from Stanford Research Institute.

The Mexican team took responsibility for assuring that the data and model of the electrical system were adequate for the analysis. The SRI team was responsible for the decision methodology. The following discussion indicates the scope and detail incorporated in the original analysis.

Figure 3.1 is a block diagram that summarizes the original model of the Mexican electrical system.

The analysis concerned the installation of the first nuclear plant in the middle 1970's. This first decision is made in the context of the overall installation and operating policy of the system.

The environment of the system is described by financial models, energy models, technology models, and electrical demand models. Given an installation and operating policy a time stream of outputs is produced by the model of the electrical system. Some of the outputs, like consumption of electricity, combine with price to yield a book profit for operation. The book profit is adjusted for quality of service as measured by outages to produce system profit.

Since the Mexican power system is a government monopoly, it is influenced by measures other than profit alone. Certain outputs other than electricity are produced by the system operation: these outputs may have either positive or negative values to Mexico. If these outputs are valued quantitatively, then a social value function that shows the social profit (or loss) generated by the system in addition to system profit can be realized. The combination of the two would be national profit. The time preference of Mexico would then be applied to this time stream of national profit to see which policy produces the highest

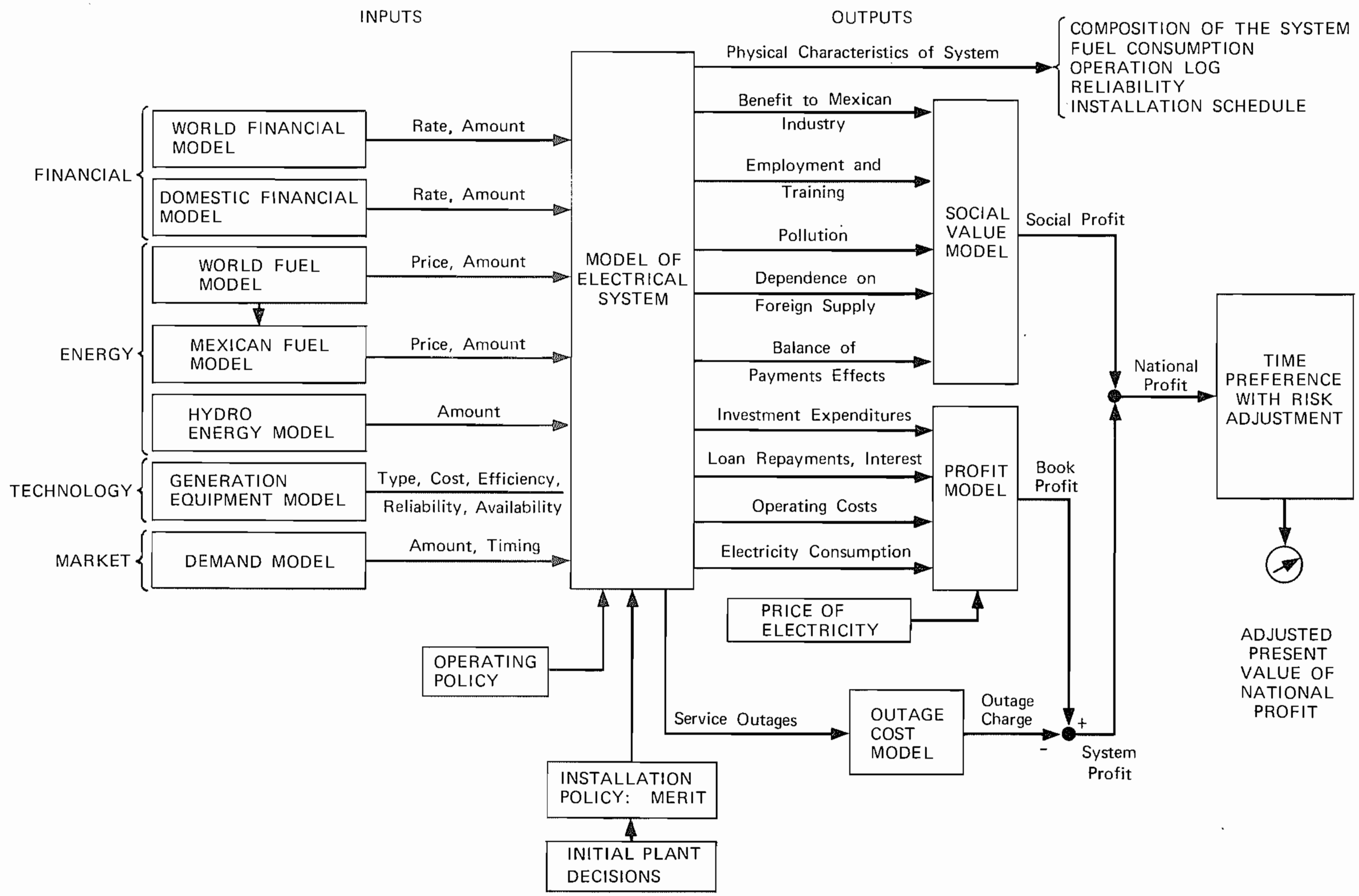


Figure 3.1: THE ORIGINAL MODEL OF THE MEXICAN ELECTRICAL SYSTEM

value. The result is a single measure of value that could be used in assessing any system policy, and in particular the decision regarding the first nuclear decision.

The original analysis required a number of approximating assumptions. The most significant assumption is the representation of the system as if it were concentrated as a single geographical point. Thus transmission effects and alternatives were not explicitly considered.

The original analysis did not explicitly treat the relationship between the demand for electricity and the quality of service and price. A forecast of demand based on predicted service and pricing policies was used. In addition, certainty was assumed for all forecasts except plant failures, stream flows, and short-term forecasts of demand.

These and other assumptions were thought to be reasonable in terms of an analysis of the first nuclear installation. The computational advances provided in this dissertation should reduce the need for such assumptions in future analyses of this type.

The main factor limiting the scope of the original analysis was the ability of the analysts to formulate, program, and solve a complex system model. The model of the electrical system in the center of Figure 3.1 had to be quite detailed in order to capture the important interactions between individual plants. It was not desirable to translate the model into a linear or nonlinear programming format because of the restrictive assumptions that would be required. The large number of alternative policies and discrete nature of the alternatives prevented the direct use of gradient search methods. Other methods also seemed to have similar difficulties.

The approach to optimization taken in the original analysis was to generate trial policies by heuristic methods and select the policy having the highest measure of value. Eventually, a policy generation routine evolved that installed plants on the basis of an approximation to the incremental value of the plant. The approximate value of a plant was computed from readily available information and parameters which can be interpreted as prices. These prices were set by an iterative process.

The intuitive optimization ideas described above were developed without a firm theoretical foundation. The lack of theoretical tools for problems of this type motivated the author's research in this area. The results of this dissertation are mechanically quite different from the methods of the original analysis but are very similar in the general approach to optimization.

3.3 Formulation of the Planning Example

In formulating the planning example it is useful to suppress many of the details of the electrical system model that do not cause difficulty in decomposing the problem. For example, social values are not explicitly treated in the planning example because it turns out that the social values treated in the original analysis can be accounted for by modifications to the parameters of the model that we will describe.[†]

[†] In situations where social values are more critical, the methods developed in Section 2.3 for problems with multiple measures of performance are applicable.

Our decision problem is to choose an installation policy for the expansion of the electrical system. By a policy we mean a complete list of plants to be installed over the planning period of the analysis. Actually, the only decision to be based on the analysis is the next installation in time; the model can be rerun before the decision on the second installation is made. Thus, some approximation in the policy for installations beyond the first installation is reasonable.

In order to describe the problem mathematically the following notation is used. Let

θ ≡ an installation policy. Given a policy $\underline{\theta}$ a complete technical description of each plant installed by the policy is available.[‡] The plants in a policy are indexed by the integers $1, \ldots, J$ where J is the maximum number of plants in a policy. Thus

$$\underline{\theta} = (\theta_1, \ldots, \theta_J)$$

where θ_j describes the j^{th} plant in policy $\underline{\theta}$.

Θ ≡ set of all possible installation policies. In Section 3.3 the properties required of this set for decomposition are discussed.

T ≅ horizon of planning period $(t = 0, \ldots, T)$.

$\pi_t(\underline{\theta})$ ≡ overall (national) profit[†] from the operation of the system in period t while under policy $\underline{\theta}$.

[†] See Figure 3.1.

[‡] Including the installation date of the plant.

$V_{T+1}(\underline{\theta}) \equiv$ terminal value of the system after horizon year T under policy $\underline{\theta}$.

The decision problem is to select a policy $\underline{\theta} \in \Theta$ to maximize the present value of profit given by

$$V(\underline{\theta}) = \sum_{t=0}^{T} \gamma_t \pi_t(\underline{\theta}) + \gamma_{T+1} V_{T+1}(\underline{\theta})$$

where the γ_t's are discount factors that reflect the time preference of the decision maker.[†]

The overall profit $\pi_t(\underline{\theta})$ is composed of a number of revenue and cost cash flows. A separate model for each cash flow is developed in this section. Given a policy, each cash flow is assumed to be independent of the other cash flows. Notationally, let

$R_t \equiv$ revenue received by the system in period t due to charges for the electrical energy delivered. The revenue is independent of the installation policy in this example.

$F_t(\underline{\theta}) \equiv$ fixed operating cost of the system in period t under policy $\underline{\theta}$. This cost includes the cost of routine maintenance, staff, and other overhead of the system.

$O_t(\underline{\theta}) \equiv$ variable operating cost of the system in period t under policy $\underline{\theta}$. This cost covers the cost of fuel and other expenses that depend on the amount of electrical energy generated.

[†] A more fundamental approach to time preference was discussed in Section 2.3. In terms of the more fundamental approach, the discount factor γ_t can be viewed as the price at which the decision maker will trade units of profit in the initial period for units of profit in period t. The assumption implied by the present value model is that the price γ_t is insensitive to the flow of profits.

$C_t(\underline{\theta}) \equiv$ reliability outage charge in period t under policy $\underline{\theta}$. this cost is a paper or actual adjustment to the books of the system to account for the quality of service as measured by outages.

$I_t(\underline{\theta}) \equiv$ installation cost of plants in period t under policy $\underline{\theta}$. This cost is the cash flow required to purchase generating equipment. The cash flow includes the effects of financing.

The overall profit in each period is simply the revenue less the sum of the costs and is given by

$$\pi_t(\underline{\theta}) = R_t - F_t(\underline{\theta}) - O_t(\underline{\theta}) - C_t(\underline{\theta}) - I_t(\underline{\theta}), \quad t = 0, \ldots, T .$$

The models underlying these functions are structured in the following subsections.

Revenue Model

Revenue from the sale of electrical energy and demand for electricity are assumed to be independent of the installation policy in this model. The purpose of including a revenue model is to retain the objective of profit maximization. Although cost minimization is mathematically equivalent in this case, the inclusion of a revenue model makes explicit assumptions that might remain hidden if cost minimization were the objective.

The major source of revenue for a power system is from the sale of electrical energy to the system's customers. Let

$\rho_t \equiv$ price of electrical energy in period t in dollars per megawatt hour (\$/MW-hr)

$\overline{d}_t \equiv$ average demand[†] in period t (MW)

$\Delta \equiv$ number of hours per period (hr).

The revenue is given by

$$R_t = \rho_t \overline{d}_t \Delta \ .$$

The price of electricity ρ_t is fictitious since the distribution system is not modeled. The price can be viewed as accounting price charged by the generating system for energy delivered to the distribution system.

The important assumption in this revenue model is that demand is independent of the installation policy. In a more detailed model we would be interested in the effect of price and the quality of service on the demand. Demand forecasting is discussed in Section 3.6 and Chapter V. The amount of revenue lost as a result of outages is accounted for in the reliability outage charge model.

Fixed Operating Cost Model

The fixed operating cost model accounts for the cost of maintenance, operating staff, and other overhead charges associated with the operation of the system. We assume that the fixed operating cost of a plant is

---

[†] The demand for electricity in industry parlance is the instantaneous rate at which energy is supplied by the system. Thus, average demand during a period multiplied by the duration of the period is equal to the total energy supplied during the period.

independent of effects attributable to other plants in the system. Hence, the fixed operating cost of the system during period t is the sum of the fixed operating costs of the plants installed in the system in period t. The fixed operating cost is given by

$$F_t(\underline{\theta}) = \sum_{j=1}^{J} f_j(t,\theta_j)$$

where

$f_j(t,\theta_j) \equiv$ fixed operating cost in period t of j^{th} plant in policy $\underline{\theta}$.

This formulation of the model permits the use of completely arbitrary functions (except for the independence between plants) to describe fixed operating cost. In general, the fixed operating cost of a plant depends both on the technology of the plant at the time of installation and on changes in the prices of materials and labor over the life of the plant. Specific examples of fixed operating cost models are developed for the numerical example in Section 3.5.

Variable Operating Cost Model

The variable operating cost model accounts for expenditures that depend on the amount of energy delivered by the system. The variable operating cost is affected by the variations in demand, the proportions of the various types and sizes of plants in the system, and the operating policy.[†]

[†] Very sophisticated models and optimization techniques have been developed for the economic operation of a power system. At the level of strategic planning of installations the model developed in this section appears to be adequate. In a full-scale analysis a more detailed model could be used to check the accuracy of this model.

In the following model the demand for electricity is characterized separately from the characterization of the cost of meeting the demand. First, we consider the demand model.

The demand (rate of energy flow) for electricity depends on the time of day, day of week, season, year, and random events such as weather. For strategic planning purposes we can represent the fluctuations in demand during a period by a load duration curve (Figure 3.2) or a demand frequency distribution (Figure 3.3).

The load duration curve gives the fraction of time (probability) that demand in a particular period is at least of a given magnitude. The demand frequency distribution is the derivative of the load duration curve and can be viewed as a probability density function. In Chapter V uncertainty is treated in more detail; at present, a frequency interpretation of probability is adequate.

At this point, we can allow the demand frequency distribution for a given period to be completely arbitrary. Thus, let

$g_t^d(d) \equiv$ frequency distribution on demand in period t where the area under $g_t^d(d)$ is unity, by definition.

Characterization of the instantaneous operating cost as a function of demand is more difficult. Figures 3.4 and 3.5 show two related characterizations of the operating cost of a hypothetical system.

Each plant in the system is represented by a vertical bar in Figure 3.4. The heights of the vertical bars are proportional to the operating cost per unit of output of the plants; short bars indicate the plants with the lowest operating cost per MWh produced.

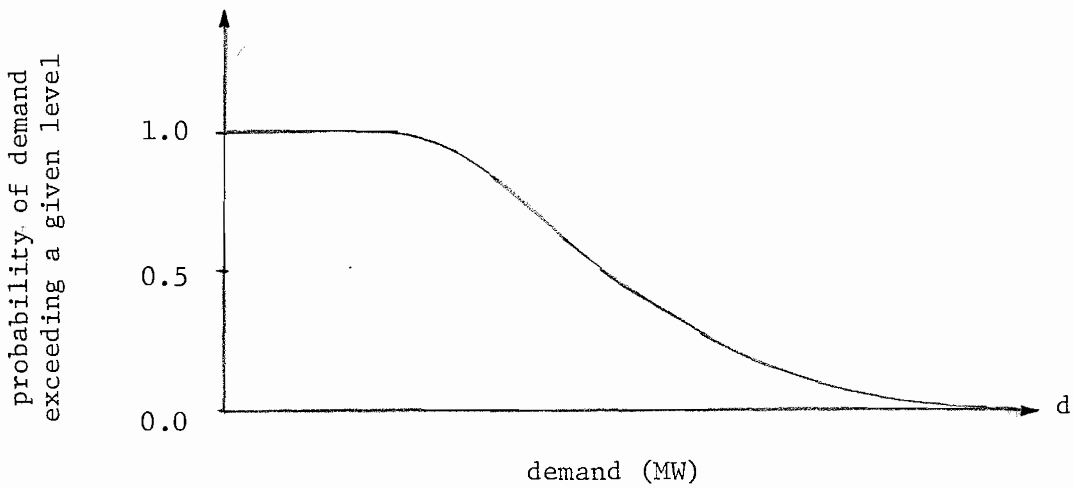


Figure 3.2: LOAD DURATION CURVE

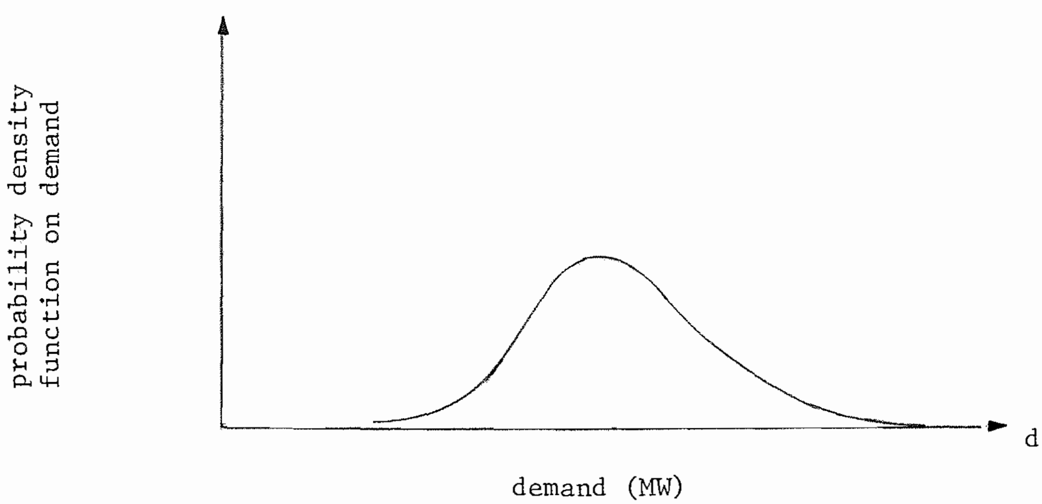


Figure 3.3: DEMAND FREQUENCY CURVE

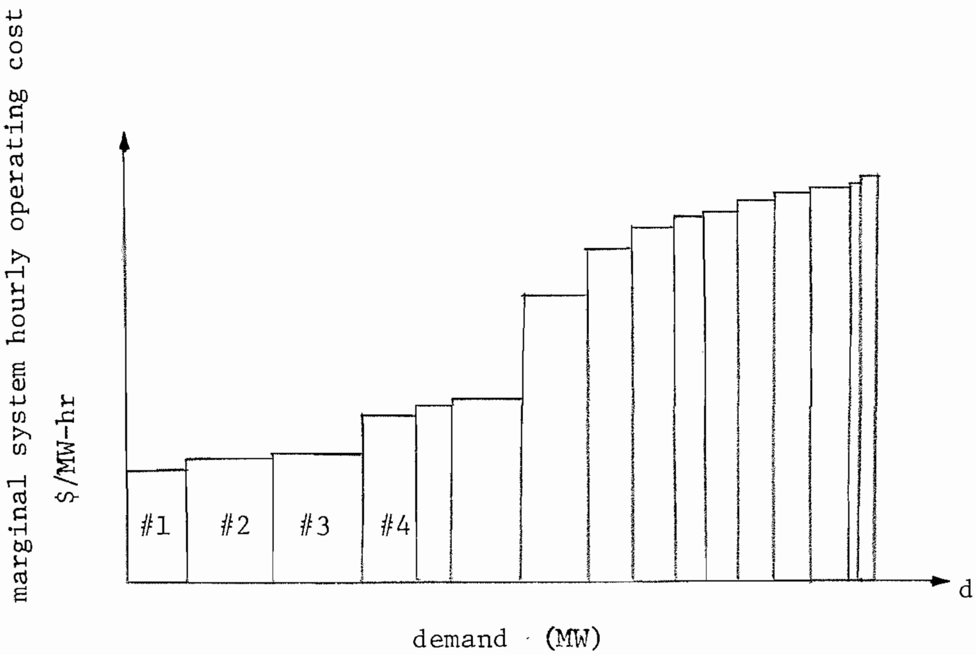


Figure 3.4: MARGINAL HOURLY OPERATING COST OF PLANTS

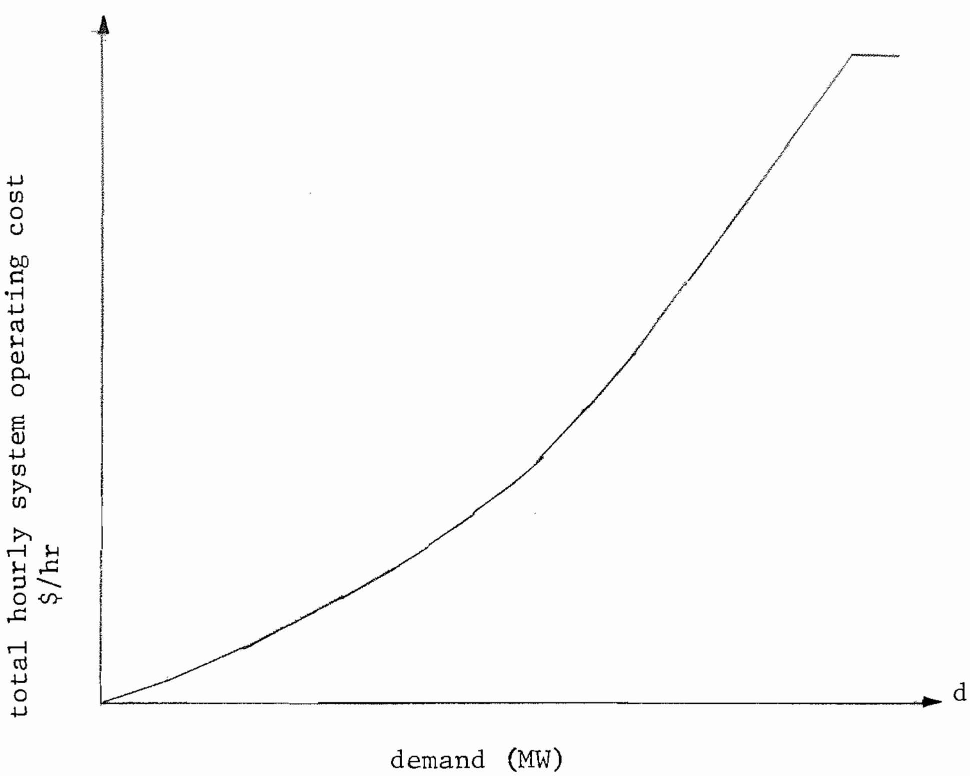


Figure 3.5: SYSTEM HOURLY OPERATING COST

For strategic planning purposes it is reasonable to assume an operating policy that loads plants in order of efficiency; the most efficient plants provide most of the energy. Thus, Figure 3.4 is also a graph of the marginal operating cost of the system (cost of satisfying an additional unit of demand).

The system hourly operating cost function is shown in Figure 3.5. This function is obtained from Figure 3.4 by integration.

A detailed description of the system hourly operating cost such as in Figure 3.5 requires knowledge of the hourly operating cost and capacity of each plant in the system. For reasons that will become clear in Section 3.4, it is necessary to characterize the system hourly operating cost in terms of a few parameters. Furthermore, these parameters should be determined by a summation of the parameters describing the independent plants in the system. The remainder of this subsection is devoted to characterizing the system hourly operating cost in this way.

In an electrical system there are four major types of plants. They are:

1. gas turbines - low capital cost but high operating cost
2. conventional thermal - medium capital cost and medium operating cost
3. nuclear - high capital cost but low operating cost
4. hydro - very high capital cost, negligible operating cost, but limited availability of sites and energy.

Hydro plants are treated as a special case in the model described below.

The variation in hourly operating cost among plants of the same type is relatively small compared to the variation in operating costs among plant types. Thus, the system hourly operating cost is adequately described by the total capacity and average hourly operating cost of each type of plant.[†]

Figures 3.6 and 3.7 show the marginal system hourly operating cost and the total system hourly operating cost curves as a function of demand where the curves are parameterized as follows:

$x_i \equiv$ amount of capacity of the j^{th} type in the system (MW).

$I \equiv$ number of types of plants in the system. Usually $I = 3$ where $i = 1,2,3$ represents nuclear, conventional thermal and gas turbine respectively (note that hydro is not included at this point).

$h_i \equiv$ total hourly operating cost of all plants of the i^{th} type ($/hr).

The advantages of this characterization of the system hourly operating cost is that the parameters x_i and h_i are easily calculated from the technical descriptions of the plants. Let

$c_{ij}(t,\theta_j) \equiv$ available capacity of the i^{th} type of plant in period t from the j^{th} plant in policy $\underline{\theta}$ (MW).

$\kappa_{ij}(t,\theta_j) \equiv$ hourly operating cost of the i^{th} type of plant in period t from the j^{th} plant in policy θ ($/hr).

[†] This assumption could be checked by comparing the results of this model with a more detailed model or results of actual system operation where available.

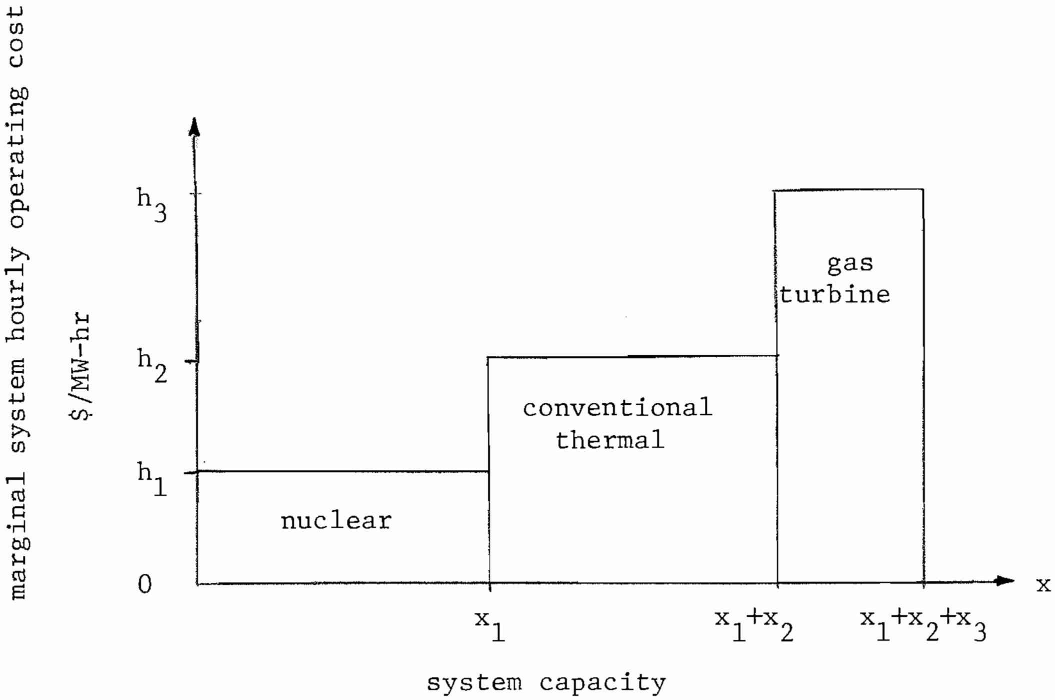


Figure 3.6: MARGINAL SYSTEM HOURLY OPERATING COST FUNCTION

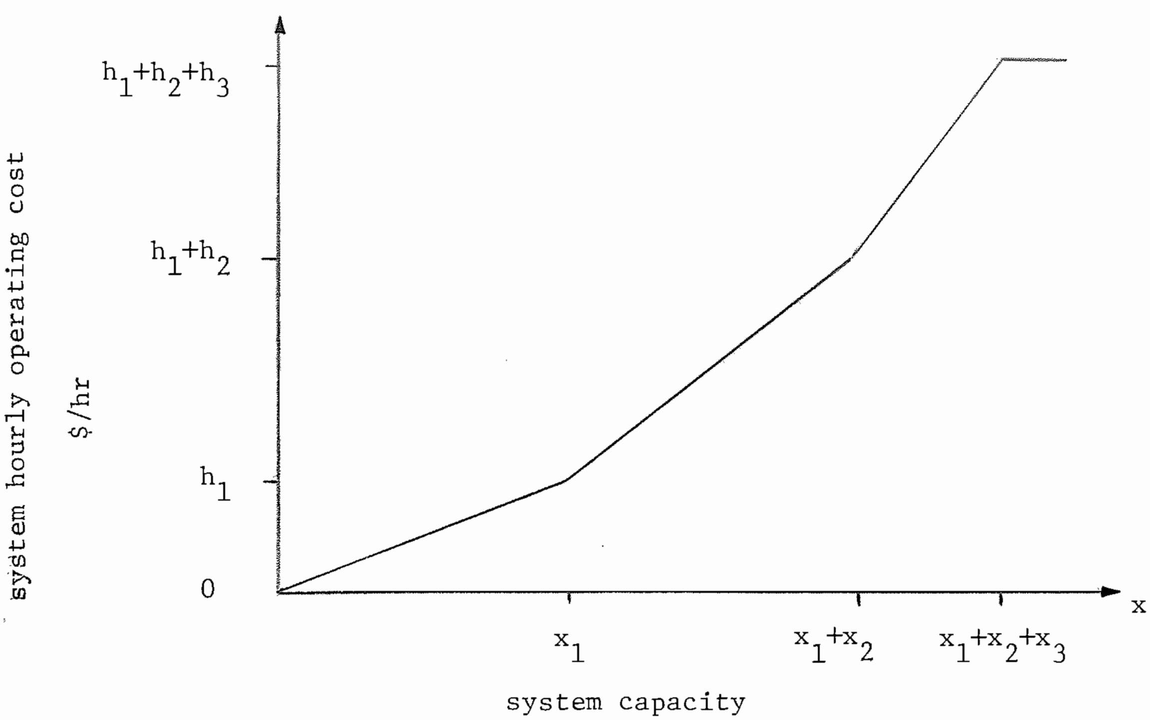


Figure 3.7: SYSTEM HOURLY OPERATING COST FUNCTION

Both of these functions are defined to be zero if the index i does not correspond to plant j's type. The dependence on t can be used to account for the effects of maintenance during a period, trends in the performance of installed plants over time, and trends in the prices of fuels over time. The parameters of the system hourly operating cost function are given by

$$x_{it} = \sum_{j=1}^{J} c_{ij}(t,\theta_j) \qquad i = 1, \dots, I, \quad t = 0, \dots, T,$$

and

$$h_{it} = \sum_{j=1}^{J} \kappa_{ij}(t,\theta_j) \qquad i = 1, \dots, I, \quad t = 0, \dots, T,$$

where a subscript has been added to the parameters of the system hourly operating cost function to indicate the period. The system hourly operating cost function can be written as

$$H_t(d|\underline{x}_t(\underline{\theta}),\underline{h}_t(\underline{\theta}))$$

where

$\underline{x}_t(\underline{\theta}) = (x_{it}(\underline{\theta}), \dots, x_{It}(\underline{\theta}))$, vector of total available capacities of each type in period t.

$\underline{h}_t(\underline{\theta}) = (h_{iy}(\underline{\theta}), \dots, h_{It}(\underline{\theta}))$, vector of total hourly operating costs of each type in period t.

In Section 3.4 we will often find it useful to write

$$\underline{x}_t(\underline{\theta}) = \sum_{j=1}^{J} \underline{c}_j(t,\theta_j) \qquad t = 0, \dots, T$$

and

$$\underline{h}_t(\underline{\theta}) = \sum_{j=1}^{J} \underline{h}_j(t,\theta_j) \qquad t = 0, \dots, T$$

where

$$\underline{c}_j(t,\theta_j) = (c_{1j}(t,\theta_j), \ldots , c_{Ij}(t,\theta_j)) \qquad j = 1, \ldots , J ,$$

$$\underline{\kappa}_j(t,\theta_j) = (\kappa_{1j}(t,\theta_j), \ldots , \kappa_{Ij}(t,\theta_j)) \qquad j = 1, \ldots , J .$$

Models of both the demand and the cost of meeting the demand have now been developed. The system variable operating cost in period t is given by

$$O_t(\underline{\theta}) = O_t(\underline{x}_t(\underline{\theta}),\underline{h}_t(\underline{\theta}))$$

where†

$$O_t(\underline{x}_t(\underline{\theta}),\underline{h}_t(\underline{\theta})) = \Delta \cdot \int_d H_t(d|\underline{x}_t(\underline{\theta}),\underline{h}_t(\underline{\theta})g_t^d(d) \ .$$

This computation essentially involves weighting the system hourly operating cost function for each demand level by the fraction of the time the system is at that demand level.

Hydro plants are treated as a special case because the total amount of energy available from a plant during a given period is limited by the available water flow and water storage capacity. We will model hydro plants as a source of energy that is used to reduce the demand placed on the thermal (non-hydro) plants.

Figures 3.8 and 3.9 illustrate a method of operating hydro plants that is reasonable for strategic planning purposes. Let

$x_0 \equiv$ amount of hydro generating capacity in the system (MW)

† The generalized integration symbol $\int$ implies summation when the variables are defined on discrete sets. Integration is implied when the sets are continuous.

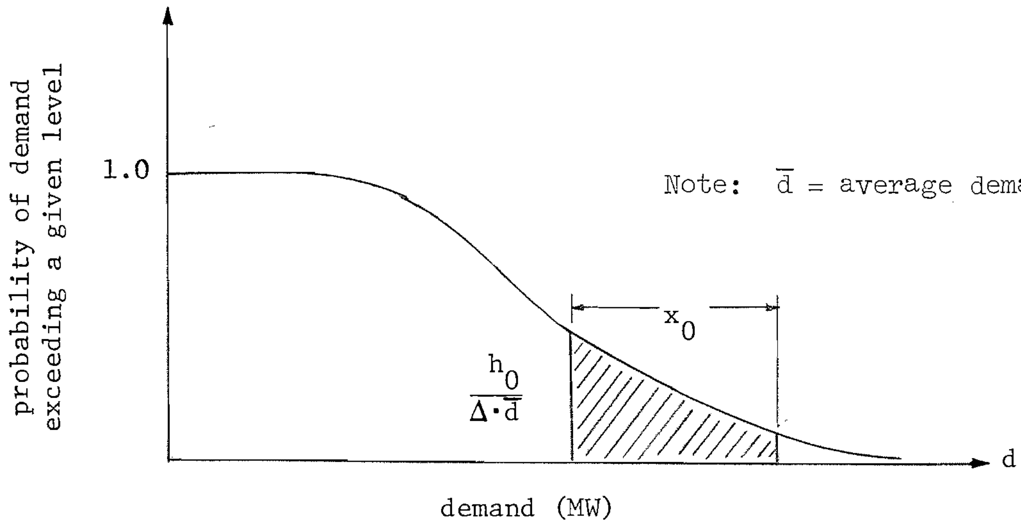


Figure 3.8: HYDRO ALLOCATION ON LOAD DURATION CURVE

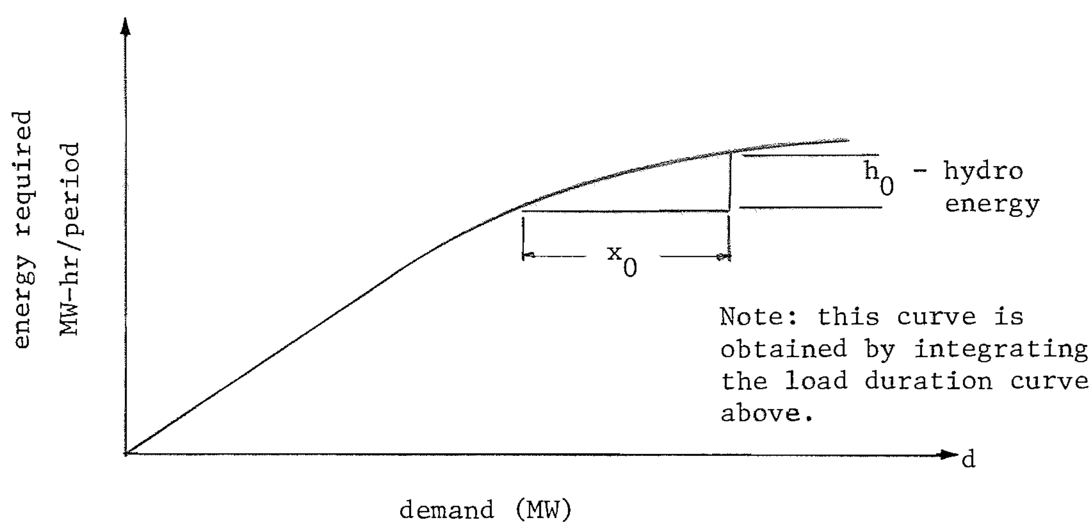


Figure 3.9: HYDRO ALLOCATION ON ENERGY CURVE

$h_0 \equiv$ total available hydro energy during a given period (MW-hr).[†]

The shaded area in Figure 3.8 represents the available hydro energy divided by the total energy requirements. The hydro energy is used most effectively when it displaces the least efficient thermal plants that operate at times of high demand.[‡] Graphically, the best operating policy for hydro plants is found by moving the indicated regions of Figures 3.8 and 3.9 as far to the right as is possible while still employing all of the hydro energy, h_0. At some point, further movement of the shaded area to the right will be limited by the available hydro capacity x_0.

The allocation of hydro energy by the method described above is easily implemented in a computer model in terms of the operation pictured in Figure 3.9. In relation to the allocation of thermal energy, the hydro energy simply removes a section from the frequency distribution on demand used to calculate the system variable operating cost.

The variable operating cost model with hydro included can be described in the same notation. The period system variable operating cost is given by

$$O_t(\underline{x}_t(\underline{\theta}), \underline{h}_t(\underline{\theta}))$$

where

$\underline{x}_t(\underline{\theta}) = (x_{0t}(\underline{\theta}), \ldots, x_{It}(\underline{\theta}))$, vector of total available capacities of each type in period t.

[†] Uncertainty in the available hydro energy due to uncertainty concerning the weather could be incorporated in this model.

[‡] The idea is to use up all of the hydro energy and at the same time have the full hydro capacity available for meeting the peak demands.

$\underline{h}_t(\underline{\theta}) = (h_{0t}(\underline{\theta}), \ldots , h_{It}(\underline{\theta}))$, vector of (1) total hourly operating cost of plant types $1, \ldots , I$; and (2) total available hydro energy in the case of type 0, in period t.

The parameters of this variable operating cost model are given as before by

$$\underline{x}_t(\underline{\theta}) = \sum_{j=1}^{J} \underline{c}_j(t,\theta_j) \qquad t = 0, \ldots , T ,$$

$$\underline{h}_t(\underline{\theta}) = \sum_{j=1}^{J} \underline{\kappa}_j(t,\theta_j) \qquad t = 0, \ldots , T ,$$

where $\underline{c}_j(t,\theta_j)$ and $\underline{\kappa}_j(t,\theta_j)$ have an additional component for hydro and $\kappa_{0j}(t,\theta_j)$ refers to available hydro energy of the j^{th} plant in period t rather than the hourly operating cost which is zero for hydro.

In certain situations the demand for energy can exceed the combined energies available from all plants in the system. Normally, reliability considerations will assure sufficient peaking capacity. Hydro plants, however, have the characteristic that their peaking capacity often exceeds their sustainable generating capacity. If sufficient energy is not available to meet demand, then an energy deficit is said to occur.

An energy deficit involves monetary costs and inconvenience to the system's customers. The situation is not as serious as outages caused by sudden plant failures; presumably, an energy deficit can be forecasted in advance. One way to account for the costs imposed by an energy deficit is to assign a price to the energy not supplied as a result of energy deficits. This price will be greater than the variable operating cost of the least efficient plants, but less than the price assigned to reliability energy losses.

In terms of the model, the cost of an energy deficit can be added to the variable system operating cost. The energy deficit is simply: The energy is not supplied after all available plants have been allocated. The cost of an energy deficit is the price assigned to a deficit times the energy not supplied.

Reliability Outage Charge Model

In this subsection a model of system reliability is developed. The model is unique because it provides an economic measure of reliability. The more usual approach to the analysis of reliability results in a technical measure of reliability such as "probability of loss of load." In an analysis of installation decisions the technical measure of reliability is often constrained to be above or below a given level. Alternatively, an economic measure of reliability avoids all the disadvantages of artificial constraints that are discussed in Section 2.1.

For strategic installation planning a reasonable economic measure of reliability is an outage charge based on the energy demanded but not supplied (energy loss) because of insufficient available generating capacity. In the original analysis a price on energy loss was determined from a previous analysis of the overall Mexican economy.

The outage charge for a given period is an uncertain quantity. In this analysis, average (expected value) outage charge is used as the economic measure of reliability. The average outage charge in period t under installation policy $\underline{\theta}$ is given by

$$c_t(\underline{\theta}) = \varphi_t \, \Delta \, \bar{\epsilon}_t$$

where

$\varphi_t \equiv$ price on energy loss (\$/MW-hr),

$\Delta \equiv$ duration of period (hrs),

$\bar{\epsilon}_t \equiv$ average capacity deficit (MW).

Thus the average energy loss in period t is $\Delta\bar{\epsilon}_t$.

At a given instant in period t,

$$\epsilon_t = d_t - c_t$$

where

$\epsilon_t \equiv$ capacity deficit (MW),

$d_t \equiv$ demand (MW),

$c_t \equiv$ available capacity (MW).

The average capacity deficit depends on the probability distribution on capacity and the demand frequency distribution according to the relation

$$\bar{\epsilon}_t(\underline{\theta}) = \int\limits_{\substack{c,d \\ d \geq c}} (d-c) g_t^c(c|\underline{\theta}) g_t^d(d)$$

where

$g_t^c(c|\underline{\theta}) \equiv$ probability distribution on available capacity in period t under installation policy $\underline{\theta}$.

$g_t^d(d) \equiv$ demand frequency distribution for period t. This distribution is identical to the demand model used for computing variable operating cost in the previous subsection.

The generalized integration required above is performed over the portion of the sample space where demand exceeds available capacity.

The probability distribution on available capacity is determined from the probability distribution on capacity of each plant in the system. Each plant in the system is assumed to be statistically independent of all other plants with regard to forced outages. Mathematical convolution can be used to obtain the system capacity distribution from the independent plant capacity distributions [9].

For decomposition we would like to parameterize the system capacity distribution so that the parameters are obtained by a summation of parameters describing the individual plants (similar to the parametrization of the system hourly operating cost in the previous subsection). Three possible parameters are obvious: the total available (nameplate) capacity, average available capacity, and variance of the available capacity can be computed by summations of the total, average and variance of the available plant capacities. Further parameters having the same mathematical characteristics (additivity for independent distributions) are given by the cumulants of the probability distributions [16].

The reliability model in the original analysis used convolution for determining the system capacity distribution. The probability distribution on available capacity of each plant is represented by a Bernoulli probability distribution where the probability of failure of a plant is defined by the ratio

$$\frac{\text{time on forced outage}}{\text{time on forced outage + time available for operation}} .$$

The failure probabilities are assigned on the basis of both historical and subjective information. The system capacity distribution is obtained from the Bernoulli probability distributions by numerical convolution with minor approximations to account for the irregular sizes of plants.

A rough empirical analysis of the results of the original analysis shows that a three parameter characterization of the system capacity distribution is adequate. The Weibull and Gamma distribution are two distributions that provide a close visual fit to the probability distributions generated in the original analysis. In order to determine the appropriate distributions for other power systems, a detailed analysis using convolution would have to be performed.

Based on the previous discussion the average outage charge is given by

$$O_t(\underline{\theta}) = O_t(x_t(\underline{\theta}), \bar{x}_t(\underline{\theta}), \check{x}_t(\underline{\theta}))$$

$$= \int_{\substack{c,d \\ d \geq c}} (d-c) g_t^c(c \mid x_t(\underline{\theta}), \bar{x}_t(\underline{\theta}), \check{x}_t(\underline{\theta})) g_t^d(d)$$

where

$x_t(\underline{\theta}) \equiv$ total installed capacity in period t under policy $\underline{\theta}$.

$\bar{x}_t(\underline{\theta}) \equiv$ average available capacity in period t under policy $\underline{\theta}$.

$\check{x}_t(\underline{\theta}) \equiv$ variance of available capacity in period t under policy $\underline{\theta}$.

The parameters $x_t(\underline{\theta})$, $\bar{x}_t(\underline{\theta})$ and $\check{x}_t(\underline{\theta})$ are computed as follows:

$$x_t(\underline{\theta}) = \sum_{j=1}^{J} c_j(t, \theta_j)$$

$$\bar{x}_t(\underline{\theta}) = \sum_{j=1}^{J} \bar{c}_j(t,\theta_j)$$

$$\check{x}_t(\underline{\theta}) = \sum_{j=1}^{J} \check{c}_j(t,\theta_j)$$

where

$c_j(t,\theta_j) \equiv$ installed capacity in period t of the j^{th} plant in policy $\underline{\theta}$.

$\bar{c}_j(t,\theta_j) \equiv$ average available capacity in period t of the j^{th} plant in policy $\underline{\theta}$.

$\check{c}_j(t,\theta_j) \equiv$ variance of available capacity in period t of the j^{th} plant in policy $\underline{\theta}$.

The functions $c_j(t,\theta_j)$, $\bar{c}_j(t,\theta_j)$ and $\check{c}_j(t,\theta_j)$ allow the probability distributions on the plant capacities to be completely general functions of time. Thus, maintenance, break-in periods of low reliability, and the effects of age can be accounted for within this model.

Another important effect that is easily incorporated into the reliability model is short-term uncertainty in demand. Short-term uncertainty in demand can be expressed in terms of a probability distribution on the parameters of the load duration curve for a given period. The time at which the probability distribution is assigned is at the last opportunity to install new capacity to be operated in the given period. The load duration curve and the probability distribution on the parameters of the load duration curve can be combined into a new load duration curve by integrating over the parameters of the load duration curve. The resulting curve is used in the model in the same way as the original curve. Generally, the new curve will have a sharper peak and a lower load factor. In Chapter VI we discuss power system planning under uncertainty in more detail.

Installation Cost Model

The installation cost model determines the cash flow resulting from the capital costs of new generating equipment. In this model the installation cost cash flows are assumed to be independent among plants. This assumption is valid if the method of financing a plant does not affect the rate or amount of financing on all other plants.

The total installation cost cash flow in period t is given by

$$I_t(\underline{\theta}) = \sum_{j=1}^{J} y_j(t,\theta_j)$$

where

$y_j(t,\theta_j) \equiv$ installation cost cash flow in period t for the j^{th} plant in policy $\underline{\theta}$.

The functions $y_j(t,\theta_j)$ include the effects of financing. Hence, these functions account for construction costs, funds borrowed, and interest and principle on debt. Within these functions, extremely general financial models are possible.

Terminal Value Model

The terminal value model assigns an approximate value to the system at the end of the planning period. A good terminal value model often greatly reduces the cost of an analysis by reducing the number of periods requiring detailed analysis.

A terminal value model is difficult to construct because a fully accurate model is as complicated as the entire system model. Clearly, only a rough approximation is reasonable. The effects of the terminal value model can be checked by sensitivity analysis.

We will approximate the terminal value of the system by assigning independent terminal values to individual plants. Thus, the terminal value of the system in period $T + 1$ under policy $\underline{\theta}$ is given by

$$V_{T+1}(\underline{\theta}) = \sum_{j=1}^{J} v_j(T+1,\theta_j)$$

where

$v_j(T+1,\theta_j) \equiv$ terminal value in period $T + 1$ assigned to the j^{th} plant in policy $\underline{\theta}$.

The terminal values might be assigned on the basis of the type, size and age of the plant at the horizon. The terminal value of the plant includes the remaining installation cost cash flows discounted to the period $T + 1$ at an appropriate discount rate.

3.4 Decomposition of the Planning Example

The planning example formulated in Section 3.3 can be stated in a form similar to that used for the multi-resource problem in Section 2.2. The decision problem is to select a policy $\underline{\theta} \in \Theta$ to maximize

$$\sum_{t=0}^{T} \gamma_t \Bigg\{ R_t - \sum_{j=1}^{J} [f_j(t,\theta_j) + y_j(t,\theta_j)]$$
$$- O_t\Big(\sum_{j=1}^{J} \underline{c}_j(t,\theta_j),\ \sum_{j=1}^{J} \underline{\kappa}_j(t,\theta_j)\Big)$$
$$- C_t\Big(\sum_{j=1}^{J} c_j(t,\theta_j),\ \sum_{j=1}^{J} \overline{c}_j(t,\theta_j),\ \sum_{j=1}^{J} \check{c}_j(t,\theta_j)\Big)\Bigg\}$$
$$+ \gamma_{T+1} \sum_{j=1}^{J} v_j(T+1,\theta_j) .$$

In this form, Theorem II in Section 2.2 applies directly to this problem.

The formulation of this power system model to the point where Theorem II can be applied is a creative process. The model is described in Section 3.3 with the benefit of hindsight. One of the useful tools in decomposing this problem was to formulate the necessary conditions and proceed as suggested in Section 2.1.

Decomposition of the problem requires that the policy set Θ be separable, i.e.,

$$\Theta = \Theta_1 \times \cdots \times \Theta_J \ .$$

In terms of the planning example, this condition requires that the availability of a plant for installation be independent of the installation of other plants. However, by restructuring the problem even this limitation can be overcome, if necessary.

Successive Approximations Algorithm

Once the problem is described in the form used above, the results of Section 2.2 can be applied (or rederived) with very little effort. The successive approximations algorithm is particularly interesting at this point because of the insight it provides. For example, an economic definition of the prices is given by Step 3 of the successive approximations algorithm.

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Estimate initial prices

$$\lambda_t^c, \qquad t = 0, \ldots , T,$$

$\lambda_t^{\bar{c}}$, $\qquad t = 0, \ldots , T,$

$\lambda_t^{\check{c}}$, $\qquad t = 0, \ldots , T,$

λ_{it}^{oc}, $\qquad t = 0, \ldots , T, \quad i = 0, \ldots , I,$

λ_{it}^{oh}, $\qquad t = 0, \ldots , T, \quad i = 0, \ldots , I,$

or start with a trial policy at Step 3.

2. Maximize

$$\sum_{t=0}^{T} \gamma_t \Big\{ -f_j(t,\theta_j) - y_j(t,\theta_j)$$
$$+ \lambda_t^c c_j(t,\theta_j) + \lambda_t^{\bar{c}} \bar{c}_j(t,\theta_j) + \lambda_t^{\check{c}} \check{c}_j(t,\theta_j)$$
$$+ \sum_{i=0}^{I} [\lambda_{it}^{oc} c_{ij}(t,\theta_j) + \lambda_{it}^{oh} \kappa_{ij}(t,\theta_j)] \Big\} + \gamma_{T+1} v_j(T+1,\theta_j)$$

over all $\theta_j \in \Theta_j$. Repeat for $j = 1, \ldots , J$. Call the results θ_j^h.

3. Calculate new prices according to the relations

$$\lambda_t^c = - \frac{\partial}{\partial x_t} C_t(x_t,\bar{x}_t,\check{x}_t) \Big|_{x_t^n,\bar{x}_t^n,\check{x}_t^n} \qquad t = 0, \ldots , T ,$$

$$\lambda_t^{\bar{c}} = - \frac{\partial}{\partial \bar{x}_t} C_t(x_t,\bar{x}_t,\check{x}_t) \Big|_{x_t^n,\bar{x}_t^n,\check{x}_t^n} \qquad t = 0, \ldots , T ,$$

$$\lambda_t^c = - \frac{\partial}{\partial x_t} C_t(x_t,\bar{x}_t,\check{x}_t) \Big|_{x_t^n,\bar{x}_t^n,\check{x}_t^n} \qquad t = 0, \ldots , T ,$$

$$\lambda_{jt}^{oc} = - \frac{\partial}{\partial x_{jt}} O_t(\underline{x}_t,\underline{h}_t) \Big|_{\underline{x}_t^n,\underline{h}_t^n} \qquad \begin{matrix} i = 0, \ldots , I \\ t = 0, \ldots , T \end{matrix}$$

$$\lambda_{it}^{oh} = - \frac{\partial}{\partial h_{it}} O_t(\underline{x}_t,\underline{h}_t) \Big|_{\underline{x}_t^n,\underline{h}_t^n} \qquad \begin{matrix} i = 0, \ldots , I \\ t = 0, \ldots , T \end{matrix}$$

where

$$x_t^n = \sum_{j=1}^{J} c_j(t,\theta_j^n) ,$$

$$\bar{x}_t^n = \sum_{j=1}^{J} \bar{c}_j(t,\theta_j^n) ,$$

$$\check{x}_t^n = \sum_{j=1}^{J} \check{c}_j(t,\theta_j^n) ,$$

$$\underline{x}_t^n = \sum_{j=1}^{J} \underline{c}_j(t,\theta_j^n) ,$$

and

$$\underline{h}_t^n = \sum_{j=1}^{J} \underline{\kappa}_j(t,\theta_j^n) , \quad t = 0, \dots , T .$$

(Note: $C_t(\ )$ and $O_t(\ )$ must be concave† and differentiable.)

4. If the new prices equal the prices determined on the previous iteration, then the conditions of Theorem II are satisfied and $\underline{\theta}^n$ is equal to $\underline{\theta}^*$, the optimal installation policy. Otherwise return to Step 2 using the new prices computed in Step 3.

The mathematical aspects of this algorithm, including the concavity and differentiability requirements are discussed later in this subsection. A relaxation coefficient can be used in Step 3 of the algorithm.

Organizational Interpretation

In this subsection the successive approximations algorithm for decomposition of the planning example is interpreted in terms of a decentralized organization. The discussion is a more precise version

† In other words $-C_t(\ )$ and $-O_t(\ )$ must be convex and differentiable.

of the introductory discussion of the electrical power system planning example in Section 1.3.

The successive approximations decomposition algorithm can be viewed as guidelines for the operation of a decentralized organization designed to plan the power system.

Each maximization in Step 2 of the algorithm can be performed by a separate plant manager. According to Step 2 the plant managers should maximize "profit." The plant manager's profit is composed of two types of cash flows. One type of cash flow results from the fixed operating costs, installation cost and terminal value of a plant. The second type of cash flow is based on the following prices and resources:

Price	Resource	Definition of Resource
λ_t^c	x_t	reliability capacity in each period
$\lambda_t^{\bar{c}}$	$\bar{x}_t$	reliability average capacity in each period
$\lambda_t^{\check{c}}$	$\check{x}_t$	reliability variance of capacity in each period
λ_{it}^{oc}	x_{it}	operating capacity of each type in each period
λ_{it}^{oh}	κ_{it}	hourly operating cost of each type in each period (hydro energy in each period for $i = 0$).

Step 2 of the algorithm directs a plant manager to choose the installation policy for his plant that maximizes the present value of profit at the given discount rates. Generally the installation decision concerns the size of the plant, but other decisions may be treated. In many cases the profit maximizing decision is not to install a plant.

The important feature of the plant manager's task is that given a set of prices the plant manager's decisions are independent of the decisions of other managers.

The prices on the resources are set in Step 3 of the algorithm. We can view these prices as the responsibility of the system managers. Through the prices on the resources, the system managers control the allocation of the resources. It is appropriate that the system managers set the prices because they have access to the technical knowledge that is required.

Two types of system managers can be defined for this problem. A reliability manager and an operating manager can be viewed as setting prices on the resources produced by the plant managers in each period. The appropriate price on the resource is the marginal value of the resources in reducing the costs of satisfying the demand. The system managers do not require detailed knowledge of the plant managers decisions. The prices can be computed on the basis of the total production of resources by the plant managers. There is no point in defining a system manager for each resource because of the detailed technical information that would need to be communicated among them.

Computational Advantages

The computational effectiveness of the decomposition method in this example is potentially enormous. For example, consider the number of policies that would require evaluation if an unsophisticated direct search were attempted. If the planning horizon requires 20 periods and 10 alternative installations of plants are possible in each year then 10^{20} evaluations must be performed. Clearly a direct approach is not economic.

In contrast, each iteration of the successive approximations algorithm requires the equivalent of about one evaluation of a policy. Thus, many iterations can be performed without approaching the high cost of a direct search. Empirical results with other problems indicate the number of iterations required is usually much less than ten [6].

Implementation of the Method

Implementation is discussed in detail in Section 3.5; however, a few comments on implementation are appropriate at this point.

There are two distinct approaches to implementing the decomposition. The first approach is to define actual functions and numerical data for the model formulated in Section 3.3. The prices in Step 3 of the algorithm can be determined from the effect of perturbations of the parameters of the reliability and operating models. The required concavity assumptions in the successive approximations algorithm can be checked by a sensitivity analysis.

The other approach to implementation of the decomposition requires a detailed model such as the computer simulation model developed for the original analysis. If the detailed model is structured along the lines of the independent submodels developed in Section 3.3, then the resources and prices defined in this section can be used to decompose the detailed model. The resulting decomposition is approximate but the model used may be more realistic.

Finally, a combination of the two approaches can be used. For example, the analyst might determine a policy by the first approach and then tune the policy on the detailed model. This approach requires fewer evaluations of the detailed model than the second approach.

The advantage of using the detailed model is that interactions that are not important enough to treat explicitly through a pricing scheme can at least be treated approximately. The disadvantage is that if these interactions are significant then the resulting policy will be approximate and the required computations will be complicated by the "noise" or random effects of these other interactions.

Price Directive Gradient Algorithm

Algorithms can be modified very easily in a well-designed computer program. Thus, the analyst might initially use his intuition and the successive approximations algorithm. If necessary, he would try more sophisticated algorithms.

The price directive gradient algorithm for the planning example is described below.

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess initial prices, λ_t^c, $\lambda_t^{\bar{c}}$, $\lambda_t^{\check{c}}$, λ_{jt}^{oc}, λ_{it}^{oh}.
2. Maximize

$$\sum_{t=0}^{T} \gamma_t \Bigg\{ -f_j(t,\theta_j) - y_j(t,\theta_j) + \lambda_t^c c_j(t,\theta_j) + \lambda_t^{\bar{c}} \bar{c}_j(t,\theta_j) + \lambda_t^{\check{c}} \check{c}_j(t,\theta_j) + \sum_{i=0}^{I} [\lambda_{jt}^{oc} c_{ij}(t,\theta_j) + \lambda_{it}^{oh} \kappa_{ij}(t,\theta_j)] \Bigg\} + \gamma_{T+1} v_j(T+1,\theta_j)$$

over all $\theta_j \in \Theta_j$. Repeat for $j = 1, \ldots, J$. Call the results θ_j^n.

3. a) Maximize

$$-\lambda_t^c x_t - \lambda_t^{\bar{c}} \bar{x}_t - \lambda_t^{\check{c}} \check{x}_t - C_t(x_t, \bar{x}_t, \check{x}_t)$$

over all x_t, $\bar{x}_t$, $\check{x}_t$. Call the results x_t^n, $\bar{x}_t^n$, $\check{x}_t^n$. Repeat for $t = 0, \ldots, T$.

b) Maximize

$$-\sum_{i=0}^{I} [\lambda_{jt}^{oc} x_{it} + \lambda_{it}^{oh} h_{it}] - O_t(\underline{x}_t, \underline{h}_t)$$

over all x_{it} and h_{it}, $i = 0, 1, \ldots, I$. Call the results x_{it}^n and h_{it}^n. Repeat for $t = 0, \ldots, T$.

4. If

$$\sum_{j=1}^{J} c_j(t, \theta_j^n) = x_t^n \qquad t = 0, \ldots, T,$$

$$\sum_{j=1}^{J} \bar{c}_j(t, \theta_j^n) = \bar{x}_t^n \qquad t = 0, \ldots, T,$$

$$\sum_{j=1}^{J} \check{c}_j(t, \theta_j^n) = \check{x}_t^n \qquad t = 0, \ldots, T,$$

$$\sum_{j=1}^{J} c_{ij}(t, \theta_j^n) = x_{it}^n \qquad t = 0, \ldots, T,$$

$$\sum_{j=1}^{J} \kappa_{ij}(t, \theta_j^n) = h_{it}^n \qquad t = 0, \ldots, T,$$

then the conditions of Theorem II are satisfied and θ_j^n, $j = 1, \ldots, J$ is the optimal policy. Otherwise, compute new prices according to

$$\lambda_t^c = \lambda_t^c - \alpha\left[x_t^n - \sum_{j=1}^{J} c_j(t,\theta_j^n)\right] \qquad t = 0, \dots , T ,$$

$$\lambda_t^{\bar{c}} = \lambda_t^{\bar{c}} - \alpha\left[\bar{x}_t^n - \sum_{j=1}^{J} \bar{c}_j(t,\theta_j^n)\right] \qquad t = 0, \dots , T ,$$

$$\lambda_t^{\check{c}} = \lambda_t^{\check{c}} - \alpha\left[\check{x}_t^n - \sum_{j=1}^{J} \check{c}_j(t,\theta_j^n)\right] \qquad t = 0, \dots , T ,$$

$$\lambda_{jt}^{oc} = \lambda_{it}^{oc} - \alpha\left[x_{it}^n - \sum_{j=1}^{J} c_{ij}(t,\theta_j^n)\right] \qquad \begin{array}{l} t = 0, \dots , T , \\ i = 0, \dots , T , \end{array}$$

$$\lambda_{it}^{oh} = \lambda_{it}^{oh} - \alpha\left[h_{it}^n - \sum_{j=1}^{J} \kappa_{ij}(t,\theta_j^n)\right] \qquad \begin{array}{l} t = 0, \dots , T , \\ i = 0, \dots , T , \end{array}$$

and return to Step 2 (the prices on the left are the new prices).

Each iteration of the price directive gradient algorithm is more difficult than each iteration of the successive approximations algorithm. Step 3 of the price directive algorithm requires the solution of two multi-variable optimization problems to determine the new prices. Generally, the solution of these multi-variable problems requires more evaluations of the reliability and operating models than determination of the prices by perturbations about the current allocations of resources.

If the operating and reliability models have further special structure then the multi-variable optimizations required by the price directive algorithm can be simplified. Conceptually it is possible to decompose these multi-variable optimization problems although it is not clear that any computational advantages would result in this case.

Bounds

Upper and lower bounds on the optimal present value of profit are useful in the practical application of the algorithms. A set of bounds can be written directly using the results of Section 2.2.

The bounds described below are valid for the price directive gradient algorithm in all cases, and for the successive approximations algorithm when the relaxation coefficient is set to unity.

BOUNDS:

$$\begin{aligned} V^{\ell} = \sum_{t=0}^{T} \gamma_t [R_t &- F_t(\underline{\theta}') - I_t(\underline{\theta}') \\ &- O_t(\underline{x}_t(\underline{\theta}'), \underline{h}_t(\underline{\theta})) - C_t(x_t(\underline{\theta}'), \bar{x}_t(\underline{\theta}'), \check{x}_t(\underline{\theta}'))] \\ &+ \gamma_{T+1} V_{T+1}(\underline{\theta}') \end{aligned}$$

and

$$\begin{aligned} V^{u} = \sum_{t=0}^{T} \gamma_t \{R_t &- F_t(\underline{\theta}') - I_t(\underline{\theta}') \\ &- O_t(\underline{x}_t', \underline{h}_t') - C_t(x_t', \bar{x}_t', \check{x}_t')] \\ &+ \lambda_t^{c}[x_t(\underline{\theta}') - x_t'] \\ &+ \lambda_t^{\bar{c}}[\bar{x}_t(\underline{\theta}) - \bar{x}_t'] \\ &+ \lambda_t^{\check{c}}[\check{x}_t(\underline{\theta}) - \check{x}_t'] \\ &+ \sum_{i=0}^{I} (\lambda_{it}^{oc}[x_{it}(\underline{\theta}') - x_{it}'] + \lambda_{it}^{oh}[h_{it}(\underline{\theta}') - h_{it}']\} \\ &+ \gamma_{T+1} V_{T+1}(\underline{\theta}') \ . \end{aligned}$$

The terms involving $\underline{\theta}'$ are determined in Step 2 of the algorithms. The terms such as x_t' are determined in Step 3 of the price directive algorithm or in Step 2 of the previous iteration in the successive approximations algorithm. An upper bound is not available on the first iteration of the successive approximations algorithm.

Gaps

It is difficult to predict whether the optimal solution to the planning example lies in a gap. Generally, a numerical example must be formulated and solved to resolve this question.

From a practical point of view, the best approach in complex strategic problems is to ignore gaps unless they cause difficulty. The upper and lower bounds on the optimal present value of profit can be used to determine the importance of a gap. If a gap is significant then the computationally less desirable penalty function methods can be used. An alternative approach for treating gaps in the planning example is developed in the next subsection.

Intuitively, gaps are most likely to involve the project managers' decisions. If the project managers' decisions oscillate between extreme alternatives for changes in the resource prices, then the danger exists that an intermediate decision is optimal. If this oscillating behavior is observed the analyst should investigate the cause of the oscillations. The investigation may reveal whether and where penalty function methods should be applied.

A Sequential Decomposition of the Example

The degree of decomposition implied by the algorithms discussed thus far is fairly complete. Each plant can be viewed as a separate entity with the only coordination between plants occurring through the iterative adjustment of the prices. Nevertheless, from an overall computational point of view it may be more effective to permit some direct coordination between certain plants. The sequential decomposition method developed in this subsection permits coordination between plants installed in the same year. The speed of convergence is improved and the effect of gaps tends to be reduced with the sequential decomposition at the price of less decomposition.

To formally describe the sequential decomposition we must redefine our decision variable notation. Let

$\theta_\tau \equiv$ a list of plants installed in period τ.

$\underline{\theta} \equiv$ an installation policy where

$$\underline{\theta} = (\theta_0, \ldots, \theta_T) .$$

$\Theta \equiv$ set of installation policies.

Thus, where we previously defined θ_j as a single plant, the term θ_τ now refers to a list of plants.

With this new notation our decision problem is to select a policy $\underline{\theta} \in \Theta$, to maximize

$$\sum_{t=0}^{T} \gamma_t \left\{ R_t - \sum_{\tau=0}^{T} [f_\tau(t,\theta_\tau) + y_\tau(t,\theta_\tau)] \right.$$
$$- C_t\left(\sum_{\tau=0}^{T} c_\tau(t,\theta_\tau), \sum_{\tau=0}^{T} \overline{c}_\tau(t,\theta_\tau), \sum_{\tau=0}^{T} \check{c}_\tau(t,\theta_\tau)\right)$$
$$\left. - O_t\left(\sum_{\tau=0}^{T} \underline{c}_\tau(t,\theta_\tau), \sum_{\tau=0}^{T} \kappa_\tau(t,\theta_\tau)\right)\right\}$$
$$+ \gamma_{T+1} \sum_{\tau=0}^{T} V_\tau(T+1,\theta_\tau) .$$

The sequential decomposition of this problem can be developed using the general method suggested in Section 2.1. Because of the notational complexity, we will only outline the development.

Differentiation of the objective function above with respect to θ_τ produces $T + 1$ simultaneous equations. Each equation will include terms of the form

$$\frac{\partial}{\partial x_t} C_t(x_t, \bar{x}_t, \check{x}_t) \Big|_{x_t^*, \bar{x}_t, \check{x}_t^*} \frac{\partial}{\partial \theta_t} c_\tau(t, \theta_\tau) \Big|_{\theta_\tau^*}$$

where

$$x_t^* = \sum_{\tau=0}^{t} c_\tau(t, \theta_\tau^*)$$

$$\bar{x}_t^* = \sum_{\tau=0}^{t} \bar{c}_\tau(t, \theta_\tau^*)$$

and

$$\check{x}_t^* = \sum_{\tau=0}^{t} \check{c}_\tau(t, \theta_\tau^*) \ .$$

We can solve the simultaneous equations iteratively if we (1) estimate certain terms, (2) solve for the optimal decisions $\underline{\theta}^*$ and (3) calculate the values of the terms we estimated. If we guess terms of the form,

$$\frac{\partial}{\partial x_t} C_t(x_t, \bar{x}_t, \check{x}_t) \Big|_{x_t^*, \bar{x}_t^*, \check{x}_t^*}$$

for $t = 0, \ldots, T$, then we will develop the same successive approximations algorithm as before. However, if we guess only those terms multiplied by terms such as

$$\frac{\partial}{\partial \theta_t} c_\tau(t, \theta_\tau)$$

where $t \neq \tau$, and do not guess similar terms where $t = \tau$ then a different form of decomposition results. In solving the necessary conditions for $\underline{\theta}$ we must solve the equations sequentially in the order $0, 1, \ldots, T$; the solution of the first $t - 1$ equations together with the terms estimated, provide enough information to solve the t^{th} equation independently of the remaining equations.

The sequential solution of the equations describing the necessary conditions can be interpreted in terms of the theory developed in Chapter II for non-differentiable functions. A formal algorithm for sequentially determining the optimal policy is provided by the following modification of the successive approximations algorithm.

SEQUENTIAL SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Estimate initial prices as in Step 1 of the successive approximations algorithm, or start with a trial policy at Step 3.
2. Maximize

$$\sum_{t=\tau}^{T} \gamma_t[-f_\tau(t,\theta_\tau) - y_\tau(t,\theta_\tau)]$$

$$+ O_\tau\Big(\sum_{t=0}^{\tau} \underline{c}_t(t,\theta_t), \sum_{t=0}^{\tau} \underline{\kappa}_\tau(t,\theta_t)\Big)$$

$$+ C_\tau\Big(\sum_{t=0}^{\tau} c_t(t,\theta_t), \sum_{t=0}^{\tau} \bar{c}_t(t,\theta_t), \sum_{t=0}^{\tau} \check{c}_t(t,\theta_t)\Big)$$

$$+ \sum_{t=\tau+1}^{T} \gamma_t\Big\{\lambda_t^c c_t(t,\theta_\tau) + \lambda_t^{\bar{c}}\bar{c}_t(t,\theta_\tau) + \lambda_t^{\check{c}}\check{c}_t(t,\theta_\tau)$$

$$+ \sum_{i=0}^{I} [\lambda_{it}^{oc} c_{it}(t,\theta_\tau) + \lambda_{it}^{oh}\kappa_{i\tau}(t,\theta_\tau)]\Big\}$$

$$+ \gamma_{T+1}V_\tau(T+1,\theta_\tau)$$

over all $\theta_\tau \in \Theta_\tau$ for $\tau = 0$. Repeat for $\tau = 1, \ldots, T$ in ascending order of the index τ. Call the results $\underline{\theta}^n$.

3. Calculate new prices as in Step 3 of the successive approximations algorithm.

4. Terminate the algorithm or return to Step 2 as in the successive approximations algorithm.

The algorithm can be justified on the basis of a theorem similar to Theorem II. Upper and lower bounds are easily developed. A sequential price directive algorithm can also be formulated.

The interpretation of the sequential decomposition in terms of a decentralized organization provides some insights. Step 2 of the algorithm suggests that the operation of the power system in each period can be viewed as the operation of separate enterprises. The manager of each enterprise has two tasks. His first task is to decide on the amount and composition of new capacity to install in his system. He receives payment for the system at the end of his period on the basis of the amount and prices of resources incorporated in the system at that time. His second task is to set the prices on the resources he receives from the previous managers at the start of his period.

The decentralized interpretation of the algorithm emphasizes that the installation decisions made in a given year are coordinated directly with the operating and reliability models for that same year rather than indirectly through prices. Prices are still used to coordinate the installation of plants in a particular year with the operating models, reliability models, and installation decisions in other years.

There are two computational advantages to the sequential algorithm.

First, convergence will tend to be faster because all of the required coordination does not rest on the iterative determination of the prices. Secondly, the sequential algorithm tends to reduce the effect of gaps. Gaps can arise because of economies of scale in purchasing new capacity. The sequential algorithm partially balances these economies of scale with the diseconomies of scale in operating the system. These computational advantages are obtained at a slight increase in the computational difficulty of Step 2 of the algorithm. In problems under uncertainty, sequential decomposition involves other restrictions.

3.5 Numerical Solution of the Planning Example

In this section we discuss the solution of a specific numerical example of our electrical power system planning problem. This example incorporates most of the significant features one would wish to treat in the analysis of capacity expansion decisions.

The example required the development of a series of computer routines. The resulting computer program illustrates how problems of this type can be organized for solution on a computer.

The data for this example is from the original analysis of the Mexican system [10]. A summary of the data is included in this section. The summary of the data indicates the amount of detail and realism that can be incorporated in an analysis of this type.

The results of this numerical example demonstrate that the decomposition methods developed in this dissertation provide a practical tool for the analysis of complex decision problems. The example is carried to the point where the convergence and general character of the results are established. Further exercising and analysis would be justified for an actual decision problem.

The Computer Program and Data Assumptions

The structure of the computer program is described by the simple flowchart in Figure 3.10. The program is designed to use both the standard and the sequential versions of the successive approximations algorithm. After some initialization calculations are performed an initial installation policy is provided to the program. This initial policy determines a flow of resources. The initial set of prices is calculated by the price routine at the resource levels determined by the initial policy. The decision routine then scans a list of possible installations for each period in the analysis. Based on the current prices of the resources, the best installation decisions are made. Finally, the current results are displayed and the program returns to the price routine to start another iteration. The use of a time-shared computer system permits the analyst to interact with the program to change parameters and terminate the run at any point.

The price routine performs the operations necessary to calculate the first derivative of the variable operating cost and reliability cost functions with respect to the resources. The differentiation is performed numerically by making small perturbations of the resources about a given level. The price routine calls on a reliability routine and a variable operating cost routine in calculating the prices.

The reliability routine is programmed to perform the calculations described in the formulation of the problem in Section 3.2. All of the

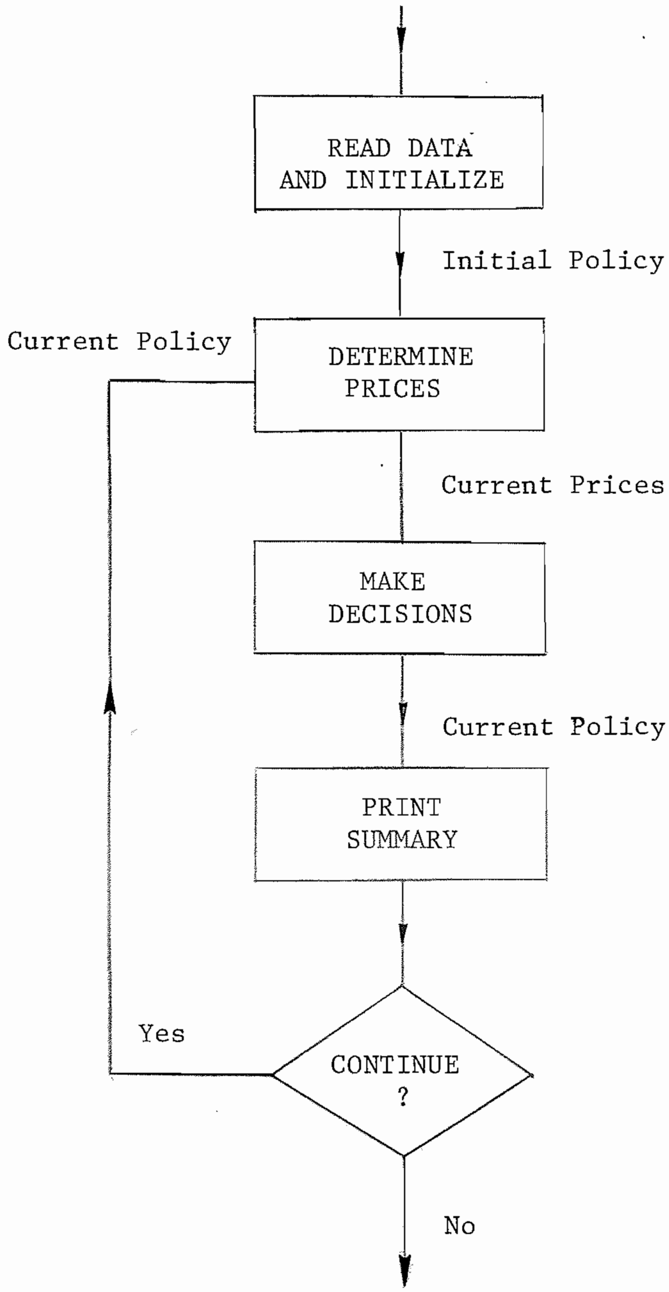


Figure 3.10: SIMPLIFIED COMPUTER FLOWCHART

required integrations are performed numerically. The reliability model requires the specification of a functional form for the probability distribution on available capacity. The results for both the Weibull and Gamma families were essentially equivalent and close to the results of the original analysis in terms of expected energy loss per year. A price of 80 U.S. cents per kWh of energy loss was used to convert expected energy loss to an economic measure of reliability.[†]

The variable operating cost model is also programmed to perform the calculations described in Section 3.3. In this example the installation of new hydro capacity is not considered because very few economic hydro sites remain undeveloped in Mexico. Thus, the hydro energy may be allocated independently of changes in the expansion policy. The program performs the allocation of hydro energy in the initialization routine of the program. Stochastic variations in hydro energy are not treated in this example, but could easily be introduced in a full-scale example. A deficit charge of 20 mills per kWh is assigned to energy deficits that occur when the energy requirements exceed the available energy from all plants in the system.

The reliability and operating models operate on an annual basis. The scheduling of maintenance and the seasonal variations in demand are not treated in this example. A first order correction for the effect of maintenance is incorporated by multiplying the available capacity of each plant by the fraction of total operating time required by

---

[†] All data in this analysis was converted to U.S. dollars at the official exchange rate of 12.5 pesos per dollar. The data is expressed in 1969 dollars.

routine maintenance. In the operating models the available capacity replaces the nameplate capacity in order to account for the effects of outages on variable operating cost.

The operating and reliability models require a load duration curve. The program stores the load duration curve in the form of a table. The load duration curve is computed from historical demand data on the Mexican system. The load factor is 0.605. The load duration curve is rescaled to account for growth in demand. The energy requirements for each year are provided as data. The growth rate is not constant but averages about 8.5 per cent per year. To reduce the number of integrations required in the operating and reliability model the integrated load curve is calculated in the initialization routine. Short-term uncertainty in demand was not incorporated in the load duration curve for this numerical example.

The decision routine generates a new policy on the basis of the current prices on the resources. The routine is designed to operate on either the standard or sequential successive approximations algorithm. At this level, the only significant difference between the two versions of the algorithm is that the sequential algorithm calls on the operating and reliability models to evaluate plants. The calculations performed by the decision routine are described in Step 2 of the algorithms. Some economies in the calculations are possible when the resources describing the plants do not vary over specified intervals of time.

In this example a plant is uniquely identified by its type, nameplate capacity, and date of installation. Plants are installed from a catalog. The catalog specifies the combinations of plants that can be installed in each year. The catalog is necessary only for the sequential algorithm, although, it can be used for the nonsequential algorithm. Table 3.1 summarizes the catalog used in the numerical example. The catalog can easily be modified to add new combinations of plants. In each period the decision routine chooses the "best" plant in the catalog.

The installation cost and fixed operating costs of the plants are summarized in Table 3.2. The installation cost is expressed as a discounted cash flow. The effects of financing, inflation and interest during construction are included in the discounted cash flow. Both the installation cost and the fixed operating cost for nuclear and thermal, display some economies of scale (decreasing average cost per MW as a function of size). The larger gas turbines are composed of a number of smaller units. Thus, the costs of gas turbines are approximately linear as a function of size. The trend in the installation costs is a decrease of 1 per cent per year for nuclear and gas turbine plants and a decrease of 0.6 per cent per year for thermal plants.

The resources or parameters describing the plants are calculated from the data in Table 3.3 according to the models described in Section 3.2. The failure probability of nuclear and thermal plants increases with size. During the break-in period of two years, the failure probability of nuclear and thermal plants is twice the normal value. In calculating the resources, the approximations for maintenance and the effect of reliability on operating cost are included. The trend in the fuel prices is a decrease of 1.7 per cent per year for nuclear and a decrease of 2.5 per cent per year for thermal and gas turbines.

Table 3.1

CATALOG OF INSTALLATION ALTERNATIVES FOR EACH YEAR

	Plant Capacity (MW)		
Catalog No.	Nuclear	Thermal	Gas Turbine
1	-	-	-
2	500	-	-
3	750	-	-
4	1000	-	-
5	-	300	-
6	-	500	-
7	-	750	-
8	-	1000	-
9	-	-	150
10	-	-	300
11	-	-	500
12	1000	-	150
13	1000	-	300
14	1000	-	500
15	-	500	150
16	-	1000	150
17	-	1000	300
18	500	-	150
19	500	-	300
20	500	500	-

Table 3.2

PLANT COST DATA

Size MW	Type	Present Value of Installation Cost Millions	Annual Fixed Operating Cost Millions Per Year
500	Nuclear	95.4	1.64
750	Nuclear	129.0	1.84
1000	Nuclear	162.6	2.04
300	Thermal	28.1	0.75
500	Thermal	38.4	0.96
750	Thermal	55.1	1.20
1000	Thermal	76.0	1.44
150	Gas Turbine	10.1	0.24
300	Gas Turbine	20.2	0.48
500	Gas Turbine	33.8	0.72

Table 3.3

PARAMETERS OF PLANT MODELS

Plant Size MW	Plant Type	Failure Probability Per Cent	Maintenance Time Months/Year	Variable Operating Cost† mills/kwh
500	Nuclear	4.0	1.0	1.35
750	Nuclear	4.8	1.0	1.30
1000	Nuclear	5.3	1.0	1.26
300	Conv. Thermal	3.0	0.5	3.13
500	Conv. Thermal	4.0	0.5	3.11
750	Conv. Thermal	4.8	0.5	3.08
1000	Conv. Thermal	5.3	0.5	2.95
150	Gas Turbine	1.0	0	4.95
300	Gas Turbine	1.0	0	4.95
300	Gas Turbine	1.0	0	4.95

† The variable operating cost is for 1969. The fossil fuel price is 33 cents per million Btu. The nominal trend in fuel prices is -1.7 per cent per year for nuclear fuel and -2.5 per cent per year for fossil fuel (conv. thermal and gas turbine).

The terminal values of the plants are assigned on the basis of their ages at the horizon and the installation cost of the plant. The terminal value model can be visualized in terms of selling the assests of the system to a hypothetical buyer at the horizon. In determining the terminal value, nuclear and thermal plants are assumed to last for 60 years and gas turbines for 40 years. The value of the plant is assumed to decrease linearly with age. For example, if a nuclear plant is 15 years old at the horizon, then 45/60th of the installation cost is assigned as the terminal value. The accuracy of this terminal value model can be tested by extending the horizon year.

The discount rate reflects the time preference of the decision makers. In this example the discount rate is 6.5 per cent per year. In assigning the prices and trends in prices the general inflation rate is factored out. In computing the effect of financing an inflation rate of 2.5 per cent per year is assumed. In actual (inflated) currency the equivalent discount rate is approximately 9 per cent.

The example assumes a six year lead time between the time a decision is made and the first operation of the plant. Assuming the first decision is made in 1969 then the first year of operation of this plant is 1975.[‡] The model is capable of simulating the operation and expansion of the system through the year 2000, or longer, if necessary, although the numerical examples were run through the year 1985.

The initial system in 1974 provides the starting point for our

---

[‡] Actually, the required lead time is considerably shorter for some plants, particularly gas turbines. In a deterministic model the length of the lead time has no effect because thereis no uncertainty to be resolved in this period.

analysis. The initial system is described by the total resources implied by the plants installed in the system in 1974. Table 3.4 summarizes the basic data from which the initial resources were calculated.

The hydro system is not affected by the decisions considered in this example. The initial peak hydro capacity is 4056 MW. The energy available from these units is 15×10^6 MWh per year.

The example was programmed on the General Electric Mark II Time-sharing Service. The budget for computer time amounted to \$1500.00 including the example in Chapter V.[†] In a full-scale analysis a reasonable computer budget would be at least an order of magnitude larger. Furthermore, several man-years of effort expended in careful gathering of data and constructing models would be reasonable in view of the magnitude of the economic resources involved in power system planning.

Results of the Numerical Example

The results of the numerical example are presented in Table 3.5. The initial policy is based on the results of the original analysis. The optimal policy was achieved in three iterations. Two additional iterations were made to demonstrate that the algorithm had converged.

The present value of the initial policy is 1192.4 million dollars versus 1194.2 million dollars for the optimal policy. Actually, only differences in the present value are significant in this example. The improvement in present value between the initial and optimal policies

---

† Each iteration of the example costs approximately \$5.00. Thus, if convergence is achieved in 6 iterations, then the cost of a complete run is approximately \$30.00. The major expense is in programming and debugging the computer program.

Table 3.4

THE MEXICAN SYSTEM IN 1974+

Number of Plants	Size MW	Type	Failure Probability Per Cent	Variable Operating Cost mills/kwh	Maintenance Months Per Year
1	27	Conv. Thermal	1.5	4.72	0.5
3	39	Conv. Thermal	1.5	3.46	0.5
2	40	Conv. Thermal	1.5	4.12	0.5
1	10	Conv. Thermal	1.5	4.72	0.5
1	5	Conv. Thermal	1.5	4.72	0.5
6	150	Conv. Thermal	3.0	3.36	0.5
2	33	Conv. Thermal	1.5	3.66	0.5
2	80	Conv. Thermal	1.5	3.66	0.5
1	300	Conv. Thermal	3.0	3.13	0.5
6	24	Gas Turbine	1.0	5.28	-
41	30	Hydro	1.0	-	-
4	52	Hydro	1.0	-	-
10	180	Hydro	1.0	-	-
4	75	Hydro	1.0	-	-
3	156	Hydro	1.0	-	-

+ Total capacity is 5755 MW in 1974.

Table 3.5

RESULTS OF THE NUMERICAL EXAMPLE

Installations from Catalog†

Iteration: Year Installed	Initial Policy	1st	2nd	3rd	4th	5th
1975	6	1	1	1	1	1
1976	2	6	9	10	10	10
1977	6	11	15	15	15	15
1978	18	15	15	15	15	15
1979	10	15	15	15	15	15
1980	13	14	15	15	15	15
1981	11	14	15	11	11	11
1982	11	14	15	15	15	15
1983	13	14	14	14	14	14
1984	11	1	11	15	15	15
1985	13	1	14	14	14	14
Present Value in Millions	1192.4	1191.3	1190.8	1194.2	1194.2	1194.2

† See Table 3.1 for definition of catalog.

is 1.8 million dollars. This difference is extremely small compared to the magnitude of the investments involved in power system planning (a 1000 MW nuclear plant costs approximately 180 million dollars with the initial fuel load).

The relatively small differences in present value among the policies generated by the algorithm occurs for three reasons. First, the initial policy was assigned on the basis of insight developed during the original analysis. In another power system, the insight of the analyst might not be as well developed. As an illustration, Table 3.6 presents the results of the algorithm starting from a less desirable policy. The algorithm does not converge as quickly as before. Nevertheless, the optimal policy is achieved.

A second reason for the relative insensitivity of the present value to the policy is important from a practical point of view. The unconstrained formulation[†] of the model balances considerations of capital cost, operating costs, reliability, timing of installations, etc., on an economic basis. In a broad region surrounding the optimal policy these considerations tend to balance out. Thus, for example, the improvement in operating costs and reliability charges resulting from early installation of a plant is approximately balanced by the effect of discounting on the earlier payment of capital costs. The practical value of the insensitivity to the policy is that other considerations not explicitly treated by the model often can be incorporated into the policy without affecting the present value index.

[†] A constraint on the amount of reserve capacity or on the probability of load loss, for example, would eliminate trade-offs between reliability and other costs.

Table 3.6

SENSITIVITY OF RESULTS TO INITIAL POLICY

Installations from Catalog[+]

Iteration: Year Installed	Initial Policy	1st	2nd	3rd	4th	5th	6th	7th	8th
1975	5	20	1	1	11	1	1	1	1
1976	2	20	6	9	6	9	10	10	10
1977	6	20	11	15	11	15	15	15	15
1978	2	20	15	15	15	15	15	15	15
1979	10	20	7	15	15	15	15	15	15
1980	3	20	7	15	15	15	15	15	15
1981	11	14	17	15	11	14	11	11	11
1982	11	14	17	15	6	11	15	15	15
1983	13	17	14	14	15	11	14	14	14
1984	11	11	1	11	7	7	15	15	15
1985	13	1	11	14	14	14	14	14	14
Present Value in Millions	1181.2	1116.2	1188.5	1190.82	1186.6	1193.4	1194.2	1194.2	1194.2

[+] See Table 3.1 for definition of catalog.

The third reason for the relative insensitivity to the policy also has practical importance. In a rapidly growing electrical system such as the Mexican system, major new installations are required at frequent intervals. The new installations provide opportunities to change the character of the system and to compensate for any undesirable effects of past decisions. The opportunity to dynamically plan the system is particularly important under uncertainty. The net effect is that uncertainty is not very important in a rapidly growing system. In Chapter V we consider power system planning under uncertainty in detail.

Some features of the optimal policy are summarized in Table 3.7. One significant feature of the policy is the relatively low reserve capacity in certain periods. In the original analysis capacity reserves on the order of 15 per cent of peak demand were found necessary. The load duration curves used in this example do not incorporate the adjustment for short-term uncertainty in demand that was suggested in the formulation of the reliability model. Since the original analysis incorporated short-term uncertainty, the difference in reserves between the two analyses is apparently the required correction for short-term uncertainty in demand. The results of the example still provide a valid demonstration of the decomposition approach. However, the model and the data must be tuned-up before the results can have policy implications for the Mexican system.

Another characteristic of the optimal policy is that nuclear plants are not installed until 1983. In the original analysis, the differences between nuclear and thermal expansion plans in the 1975-1980 period

Table 3.7

FEATURES OF THE OPTIMAL POLICY

Year	Peak Demand MW	Installed Capacity MW	Reserve Capacity Per Cent	Hydro Capacity MW	Nuclear Capacity MW	Thermal Capacity MW	Gas Turbine Capacity MW
1975	5,297	5,755	9	4056		1615	84
1976	5,761	6,055	5	4056		1615	384
1977	6,233	6,705	7.5	4056		2115	534
1978	6,758	7,355	9	4056		2615	684
1979	7,325	8,005	9	4056		3115	834
1980	7,936	6,655	9	4056		3615	984
1981	8,600	9,155	6.5	4056		3615	1484
1982	9,311	9,805	5	4056		4115	1634
1983	10,084	11,305	12	4056	1000	4115	2134
1984	10,920	11,955	9.5	4056	1000	4615	2284
1985	11,826	13,455	13.5	4056	2000	4615	2784

were small. The fact that no nuclear plants were installed in the optimal policy in this example may be attributed to small differences in the models, such as the neglect of uncertainty in hydro energy. Nevertheless, the effect is small as is evidenced by the small difference between the present value of the initial policy in Table 3.5 that installs nuclear capacity in 1976 and the present value of the optimal policy that installs no nuclear capacity until 1983.

It is interesting to demonstrate the algorithm in a situation that is favorable to nuclear power. Table 3.8 presents the results of an example where the trend in nuclear fuel price is a decrease of 10 per cent per year (from 1969) rather than the nominal decrease of 1.7 per cent per year used in the previous examples. Although this example is extreme, the results are intuitive. The best strategy is to install nuclear capacity as quickly as possible in order to achieve the operating cost savings.

We can obtain further insight by examining the prices on the resources produced by a policy. Table 3.9 contains the prices assigned to each resource produced by the optimal policy in Table 3.5. It is sometimes difficult to interpret the prices because each price is determined by the interaction of almost 100 resources! Furthermore, the magnitude of the resources depend on the units used to describe the resources. Nevertheless, the pattern of prices over time provides some insight.

The first three resources referred to in Table 3.9 concern the reliability of the system. The most important of these is the capacity resource. The price on capacity hovers around 7.0×10^{-3} except in

Table 3.8

SENSITIVITY TO NUCLEAR FUEL PRICE TREND

Installations from Catalog[+]

Iteration: Year Installed	Initial Policy	1st	2nd	3rd	4th	5th	6th	7th
1975	6	14	1	1	4	4	4	4
1976	2	14	10	9	4	4	4	4
1977	6	14	15	3	1	1	1	1
1978	18	14	15	12	3	3	3	3
1979	10	14	14	4	9	9	9	9
1980	13	14	14	3	3	3	3	3
1981	18	14	14	11	11	11	11	11
1982	10	14	14	4	13	13	13	13
1983	13	14	14	19	11	11	11	11
1984	11	11	1	19	13	14	14	14
1985	13	4	3	3	3	3	3	3
Present Value in Millions	1223.9	1130.4	1244.9	1283.6	1306.3	1306.5	1306.5	1306.5

[+] See Table 3.1 for definition of catalog.

Table 3.9

PRICES ON THE RESOURCES OF THE OPTIMAL POLICY

Resource: Year	Reliability Capacity	Average Capacity	Variance of Capacity	Nuclear Capacity	Nuclear Operating Cost	Thermal Capacity	Thermal Operating Cost	Gas Turbine Capacity	Gas Turbine Operating Cost
	$x10^{-3}$	$x10^{-3}$	$x10^{-5}$	$x10^{-2}$	$x10^{-3}$	$x10^{-2}$	$x10^{+3}$	$x10^{-2}$	$x10^{+3}$
1975	2.08	2.71	-6.70	2.59	-7.87	2.46	-7.16	0.0	0.0
1976	7.88	-56.02	-12.13	22.19	-137.18	4.73	-7.34	1.61	-3.21
1977	7.21	3.58	-10.39	2.52	-8.23	2.42	-7.13	0.006	-118.88
1978	7.34	4.40	-9.23	2.51	-8.76	2.22	-6.58	0.008	-174.81
1979	6.74	4.10	-7.83	2.60	-9.71	2.09	-6.20	0.013	-264.05
1980	7.14	3.33	-7.57	2.89	-12.30	2.01	-5.98	0.023	-467.75
1981	9.28	-25.05	-8.87	6.30	-36.88	2.26	-6.41	0.091	-18.39
1982	13.83	-5.53	-10.01	7.74	-47.86	2.21	-6.24	0.113	-22.68
1983	7.27	5.17	-5.30	2.20	-7.29	1.94	-5.49	0.070	-14.11
1984	11.25	1.45	-6.51	2.19	-7.39	1.95	-5.53	0.098	-19.86
1985	5.49	4.37	-3.59	2.08	-6.94	1.75	-4.93	0.075	-15.31

the years 1975, 1982, and 1984. In 1975 the price is relatively low indicating an excess of capacity. In 1982 and 1984 the price is relatively high indicating a slight shortage of capacity.

The prices on average capacity and variance of capacity are more difficult to intuit. Generally, the variance of capacity is an undesirable resource and has a negative price. Average capacity is usually a desirable resource with a positive price. In 1976, 1981, and 1982, the years with the smallest reserve capacity, the average capacity has a negative price. On the surface this appears strange. However, the reliability resources interact in a very complicated manner which is difficult to describe in a simple way.

The remaining six prices are assigned to the operating resources: two for each type of plant. The prices assigned to operating capacity are positive while the prices assigned to total hourly operating cost are negative. The price assigned to gas turbine capacity is relatively small, since gas turbines are primarily installed for reliability. The prices or nuclear capacity generally increase over time until nuclear capacity is actually installed in 1983.

A relaxation coefficient of 0.5 was used in all of the examples. Unfortunately, the budget for computer time did not permit a detailed investigation of the effects of changing the relaxation coefficient. The algorithm did not appear to converge when the relaxation coefficient was set to unity.

Once the algorithm had converged on a policy the relaxation coefficient was set to 1.0 for the final iteration. The conditions of the optimality theorem (Theorem II) are not formally satisfied unless the prices on successive iterations are identical. The use of a relaxation coefficient prevented the easy calculation of upper bounds on the present value.

The results of this example are optimal only if the operating and reliability models are concave. This assumption was not verified quantitatively. Intuitively there is no reason to suspect the condition is not satisfied, at least in the region of the optimal policy. In a full-scale analysis the price directive gradient algorithm could be implemented to check this assumption. The price directive algorithm does require concavity of the operating and reliability models. An advantage of the price directive algorithm is that upper bounds could be easily computed for this example. A disadvantage is that the price directive algorithm is more expensive to implement and operate.

3.6 Conclusions Based on the Model

The results of this numerical example of electrical power system planning clearly demonstrate the practical value of the methodology developed in this dissertation. Naturally, the decision to apply the methodology to another power system problem should receive careful consideration. On the negative side, the analyst is not, apriori, guaranteed that the methodology will solve his problem. Gaps and convergence difficulties may present overwhelming problems in some power system problems.

On the positive side, the methodology requires no restrictive assumptions of the model. If the methodology is used and a solution cannot be obtained because of gaps or convergence difficulties, then the model of the system is still useful. Heuristic or other optimization techniques could be applied at this point. When gaps and convergence are _not_ problems, then the methodology is far more powerful than any other approach.

3.7 Possible Extensions of the Model

We have already mentioned several areas where assumptions in the electrical power system model can be relaxed without affecting the method of solution or the structure of the model. In this subsection we discuss alternative ways of formulating electrical power system problems. A full-scale application of the methods of this dissertation to a power system planning problem would afford the opportunity to reconsider the scope of the model.

An interesting and possibly valuable extension of the model is to relax the assumption of a point system. Relaxation of this assumption would allow explicit consideration of transmission, system stability, and area protection effects. In terms of the decomposition approach, the system could be represented as several geographically separate, but interconnected, systems. The separate systems could be viewed as buying and selling power to each other. By carefully structuring the model describing the whole interconnected system, it should be possible to apply the methods of this dissertation. The result might be a set of prices that would coordinate the allocation of resources among the interconnected systems. Without first formulating such a model it is difficult to sepeculate on the exact form the prices, resources, and resource markets might take.

Another extension along the same lines is the explicit treatment of plant siting effects. In addition to the transmission, system stability, and area protection effects, plant siting is important in terms of local economic and environmental conditions. For example, a full-scale model might take into account the transportation costs of fuel, or the availability of cooling water at a particular location. The social benefits and costs of the ecological effects of power plants in various locations could be considered. This extension of the model would be particularly important for power systems in industrially advanced nations, where the environment is an important consideration.

At a higher level, the methodology can be applied to the planning of a decentralized, nation-wide power system. In the United States, the ultimate responsibility for the operation of power systems rests with the regulatory agencies. These agencies are sometimes viewed as delegating certain decisions to private and public power system managers. The decisions of the power system managers are subject to guidelines set by the regulatory agencies.

If a model of the nationwide system could be structured (not solved numerically) then it may be possible to identify methods for decomposing the model. The decomposition of the model would suggest ways of decentralizing and regulating nation-wide power systems.

CHAPTER IV

DECOMPOSITION UNDER UNCERTAINTY

The mathematical fundamentals of decomposition under uncertainty are developed in this chapter. The general outline of this chapter is similar to Chapter II where the mathematical fundamentals are developed for deterministic problems. It is significant that the decomposition of problems under uncertainty requires exactly the same mathematical tools as the decomposition of deterministic problems.

Problems with separable objective functions are treated in Section 4.1. This class of problems includes problems where the objective is to maximize the expected present value of profit.

The analysis is extended in Section 4.2 to problems with arbitrary objective functions. In this class of problems are problems where a multi-attribute risk preference function (a von Neuman-Morgenstern utility function) describes the preferences of the decision maker.

In our development of the mathematical foundations of decomposition we shall classify problems as either open-loop or closed-loop decision problems. In an open-loop decision problem the decision maker must irrevocably allocate his resources before the uncertainty is revealed. In a closed-loop decision problem the decision maker has the opportunity to adjust the allocation of his resources depending on how the uncertainty is resolved. In some cases the uncertainty is slowly resolved over time and the decision maker can respond dynamically to the new information. The extreme case of perfect information is where all the uncertainty is resolved before a decision is made. Another extreme case of closed-loop decision making is the open-loop case described above where none of the uncertainty is resolved before a decision is made.

In this chapter, the mathematical foundations of decomposition are developed for the general closed-loop decision problem which includes all other classes of decision problems as special cases. For some extreme cases of closed-loop decision making the results are relatively easy to implement. In the most general cases, specialized computational techniques or approximation methods are useful. Computational methods for decomposition under uncertainty are discussed in Section 4.3.

In passing we should note that the introduction of uncertainty greatly increases the computational difficulties associated with optimization. The results of this chapter permit more generality than is often required in practical problems. Uncertainty should be treated explicitly only where sensitivity analysis indicates that an important decision is sensitive to changes in a state variable.

4.1 Problems under Uncertainty with Separable Objective Functions

In this chapter problems involving both time and uncertainty are considered.

The treatment of time in this section builds on the results of Section 2.2. These results are applied to the electrical power system problem involving time in Chapter III. Problems involving time are treated by considering flows of resources where the flow of a resource in each discrete period is viewed as a separate resource. Resources are similarly defined in this chapter in terms of their physical characteristics and the time period.

The notation used in this chapter is extremely general. Few problems require the level of generality implied by the notation. However, the generality of the notation actually reduces the notational problems because many special cases that do not prevent decomposition are not explicitly recognized by the notation.

The Example

The example concerns the selection of a policy to maximize the expected present value of profit associated with a very general resource allocation problem under uncertainty. The notation is a straightforward extension of the notation developed in Section 2.2 for problems under certainty.

Let

$\underline{s}_t$ $\equiv$ the vector of uncertain state variables whose uncertainty is usually viewed as being resolved at the end of period t. The components of $\underline{s}_t$ do not need to be defined at this point.

$\underline{s}$ $\equiv$ the matrix of uncertain state variables where

$$\underline{s} = (\underline{s}_0, \ldots , \underline{s}_T) .$$

$\theta_{tj}(\underline{s})$ $\equiv$ the decision variable associated with the j^{th} project in period t. For example, θ_{tj} is usually viewed as being set at the start of period t when the state variables $\underline{s}_0, \ldots , \underline{s}_{t-1}$, have been resolved and $\underline{s}_t, \underline{s}_{t+1}, \ldots , \underline{s}_T$ are still uncertain. However, the notation is intended to allow θ_{tj} to depend on any subset of the uncertain state variables that are resolved when θ_{tj} is set.

$\underline{\theta}_j(\underline{s})$ $\equiv$ the vector of decision variables (a policy) associated with the j^{th} project where

$$\underline{\theta}_j(\underline{s}) = (\theta_{0j}(\underline{s}), \dots, \theta_{Tj}(\underline{s})) .$$

$\underline{\theta}(\underline{s})$ $\equiv$ the matrix of all decision variables (a policy) where

$$\underline{\theta}(\underline{s}) = (\underline{\theta}_1(\underline{s}), \dots, \underline{\theta}_J(\underline{s})) .$$

$x_{tjk}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ amount of the k^{th} resource used by the j^{th} project in period t as a function of the policy and the uncertain state variables. This function is given by a detailed structural model of the project. For example, x_{tjk} might be a function of $\underline{\theta}_0(\underline{s}), \dots, \underline{\theta}_t(\underline{s})$, and $\underline{s}_0, \dots, \underline{s}_t$, but not a function of $\underline{\theta}_{t+1}(\underline{s}), \dots, \underline{\theta}_T(\underline{s})$ and $\underline{s}_{t+1}, \dots, \underline{s}_T$. The resource x_{tjk} is assumed to flow from the resource market to the j^{th} project at the end of period t.

$\underline{x}_{tj}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the vector of resources used by the j^{th} project in period t where

$$\underline{x}_{tj}(\underline{\theta}(\underline{s}),\underline{s}) = (x_{tj1}(\underline{\theta}(\underline{s}),\underline{s}), \dots, x_{tjK}(\underline{\theta}(\underline{s}),\underline{s})) .$$

$\underline{x}_{tk}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the vector describing the allocation of the k^{th} resource in period t among the J projects where

$$\underline{x}_{tk}(\underline{\theta}(\underline{s}),\underline{s}) = (x_{t1k}(\underline{\theta}(\underline{s}),\underline{s}), \dots, x_{tJk}(\underline{\theta}(\underline{s}),\underline{s})) .$$

$\underline{x}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the 3-dimensional matrix of resource allocations.

$y_{tk}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ total amount of the k^{th} resource used in period t by all J projects. Thus,

$$y_{tk}(\underline{\theta}(\underline{s}),\underline{s}) = \sum_{j=1}^{J} x_{tjk}(\underline{\theta}(\underline{s}),\underline{s}) \; .$$

$\underline{y}_t(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the vector of the total amounts of resources employed in period t by all J projects where

$$\underline{y}_t(\underline{\theta}(\underline{s}),\underline{s}) = \left(y_{t1}(\underline{\theta}(\underline{s}),\underline{s}), \; \ldots \; , \; y_{tK}(\underline{\theta}(\underline{s}),\underline{s})\right).$$

$\underline{y}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the matrix of the total amounts of resources employed by all projects.

$R_t(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ total revenue in period t from all J projects as a function of the policy $\underline{\theta}(\underline{s})$ and the uncertain state variables $\underline{s}$. Generally,

$$R_t(\underline{\theta}(\underline{s}),\underline{s}) = \sum_{j=1}^{J} r_{tj}(\underline{\theta}(\underline{s}),\underline{s})$$

where $r_{tj}(\underline{\theta}(\underline{s}),\underline{s})$ is the revenue attributed to the j^{th} project in period t.

$C_t(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})$ $\equiv$ total cost in year t of all resources purchased in the resource markets. Generally,

$$C_t(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s}) = \sum_{k=1}^{K} C_{tk}(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})$$

where C_{tk} is the cost attributed to the k^{th} resource in period t.

The example problem is to choose a policy to maximize the expected present value of profit. Thus, we maximize

$$\int_{\underline{s}} \sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}(\underline{s}),\underline{s}) - C_t(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})] \{\underline{s}|\mathscr{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})^{\dagger}$ where

γ_t $\equiv$ the discount factor associated with period t,

$\{\underline{s}|\mathscr{E}\}$ $\equiv$ joint probability distribution on $\underline{s}$ assigned on the basis of the decision maker's prior information at $t = 0$. $\{\underline{s}|\mathscr{E}\}$ generally describes the environment of the problem and is assumed to be independent of $\underline{\theta}(\underline{s})$,

and $\Theta(\underline{s})$ $\equiv$ set of all available policies (explained below).

The concept of a policy embodied in the notation $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$ is critical to understanding what follows in this chapter. The decision variables $\theta_{tj}(\underline{s})$, comprising the policy, control the allocations of resources. The notation $\theta_{tj}(\underline{s})$ is a convenient way of describing the dependence of the decision variables on the uncertain state variables. This dependence arises only in closed-loop policies where decision variables can be set after the uncertainty in some of the state variables has been resolved. The notation $\theta_{tj}(\underline{s})$ does not imply that θ_{tj} depends on every component of $\underline{s}$. The set $\Theta(\underline{s})$ defines the uncertain state variables that are resolved at the time θ_{tj} is set

For example, consider the following three types of policies that

---

$\dagger$ The generalized integration symbol $\int$ implies summation when the probability distribution is discrete. Integration is implied when the probability distribution is continuous.

are special cases of closed-loop policies:

1. Delayed resolution policy (open-loop policy)

2. Dynamic resolution policy (adaptive policy)

3. Immediate resolution policy (perfect information policy).

In a delayed resolution policy, the decision maker sets all of the decision variables before any of the uncertain state variables are resolved. In this case, $\underline{\theta}$ does not depend on any component of $\underline{s}$. Of course, $\underline{\theta}$ still depends on the joint probability distribution on $\underline{s}$, $\{\underline{s}|\mathcal{E}\}$, through the optimization. The notation $\Theta(\underline{s})$ defines the available policies in a problem. If the analyst wishes to consider only delayed resolution policies, then $\Theta(\underline{s})$ is used to restrict consideration to policies that do not depend on $\underline{s}$.

The other extreme case concerns the immediate resolution policy where all of the decision variables are set after the uncertain state variables are resolved. In this case, $\underline{\theta}$ depends on all components of $\underline{s}$. This case is equivalent to parametrically solving the deterministic version of the problem as a function of $\underline{s}$. The set $\Theta(\underline{s})$, in this case, includes only the policies that depend on all components of $\underline{s}$.

The dynamic resolution policy is the most interesting case. In this case, the decision variables are set on the basis of the decision maker's state of information at the time he makes a commitment. The decision maker's state of information depends on the structure of the information flows in the problem. The set $\Theta(\underline{s})$ specifies the structure of these information flows by defining the state variables that are resolved at the time a decision is made. For example, $\Theta(\underline{s})$ is simply a concise way of stating that†

$$\theta_{tj}(\underline{s}) = \theta_{tj}(\underline{s}_0, \cdots, \underline{s}_{t-1}) .$$

In this example, θ_{tj} depends on the conditional distribution,

$$\{\underline{s}_t, \cdots, \underline{s}_T | \underline{s}_0, \cdots, \underline{s}_{t-1}, \mathcal{E}\} ,$$

through the optimization.‡

Mathematical Results (Theorem IV, Bounds, and Algorithms)

The following theorem provides sufficient conditions for the optimality of a solution to the example formulated in the previous subsection. The theorem is analogous to Theorems I and II in Chapter 2.

THEOREM IV: If $\underline{\theta}^*(\underline{s})$ maximizes

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \left[\gamma_t \; R_t(\underline{\theta}(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s})\right]\right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$, and if $\underline{y}^*$ maximizes

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \sum_{\tau=0}^{T}\left[\sum_{k=1}^{K} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{s}) - C_t(\underline{y}(\underline{s}),\underline{s})\right]\right] \{\underline{s}|\mathcal{E}\}$$

† The subscripts on $\underline{\theta}_t$ and $\underline{s}_t$ are useful for describing the problem structure. However, only $\Theta(\underline{s})$ defines the structure of the information flows. In the dynamic resolution case, $\Theta(\underline{s})$ often defines the same information flow structure as is implied by the subscripts on $\underline{\theta}_t$ and $\underline{s}_t$.

‡ The value of the notation developed in this chapter becomes evident when we try to write a dynamic resolution problem using more conventional notation, i.e.,

$$\max_{\underline{\theta}_0} \int_{\underline{s}_0} [\max_{\underline{\theta}_1} \int_{\underline{s}_1} [\max_{\underline{\theta}_2} \int_{\underline{s}_2} [V(\underline{\theta},\underline{s})\{\underline{s}_2|\underline{s}_1,\underline{s}_0,\mathcal{E}\}]\{\underline{s}_1|\underline{s}_0,\mathcal{E}\}]\{\underline{s}_0|\mathcal{E}\}] .$$

over all $y(\underline{s})$, and if

$$\underline{y}^*(\underline{s}) = \underline{y}(\underline{\theta}^*(\underline{s}),\underline{s})$$

for all logically possible values of $\underline{s}$, then $\underline{\theta}^*(\underline{s})$ maximizes

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t[R_t(\underline{\theta}(\underline{s}),\underline{s}) - C_t(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})]\right]\{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$.

Proof:

a) Interchanging the order of integration and summation in the two inequalities implied by the conditions of the theorem results in the following two inequalities:

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t R_t(\underline{\theta}(\underline{s}),\underline{s})\right]\{\underline{s}|\mathcal{E}\} - \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t\left[\sum_{k=1}^{K} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s})\right]\right]\{\underline{s}|\mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t R_t(\underline{\theta}^*(\underline{s}),\underline{s})\right]\{\underline{s}|\mathcal{E}\} - \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t\left[\sum_{k=1}^{K} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{\theta}^*(\underline{s}),\underline{s})\right]\right]\{\underline{s}|\mathcal{E}\}$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$, and

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t\left[\sum_{k=1}^{K} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{s})\right]\right]\{\underline{s}|\mathcal{E}\} - \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t C_t(\underline{y}(\underline{s}),\underline{s})\right]\{\underline{s}|\mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t\left[\sum_{k=1}^{K} \lambda_{tk\tau}(\underline{s}) y^*_{\tau k}(\underline{s})\right]\right]\{\underline{s}|\mathcal{E}\} - \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t C_t(\underline{y}^*(\underline{s}),\underline{s})\right]\{\underline{s}|\mathcal{E}\}$$

holds for all $\underline{y}(\underline{s})$.

b) Combining the two inequalities in Step a) gives

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}(\underline{s}),\underline{s}) - C_t(\underline{y}(\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$+ \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \lambda_{tk\tau}(\underline{s}) [y_{\tau k}(\underline{s}) - y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}^*(\underline{s}),\underline{s}) + C_t(\underline{y}^*(\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$+ \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \lambda_{tk\tau}(\underline{s}) [y^*_{\tau k}(\underline{s}) - y_{\tau k}(\underline{\theta}^*(\underline{s}),\underline{s})] \right] \{s | \mathcal{E}\}$$

which holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$ and all $\underline{y}(\underline{s})$.

c) Since the inequality in Step b) holds for all $\underline{y}(\underline{s})$ it must also hold for $\underline{y}(\underline{s}) = \underline{y}(\underline{\theta}(\underline{s}),\underline{s})$ where $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. In this case the terms involving λ on the left side of the inequality in Step b) cancel, and

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}(\underline{s}),\underline{s}) - C_t(\underline{y}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}^*(\underline{s}),\underline{s}) + C_t(\underline{y}^*(\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$+ \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \lambda_{tk\tau}(\underline{s}) [y^*_{\tau k}(\underline{s}) - y_{\tau k}(\underline{\theta}^*(\underline{s}),\underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$.

d) By the statement of the theorem

$$\underline{y}^*(\underline{s}) = y(\underline{\theta}^*(\underline{s}),\underline{s})$$

for all logically possible values of $\underline{s}$. Thus, the terms involving λ on the right side of the inequality in Step c) cancel, and

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}(\underline{s}), \underline{s}) - C_t(\underline{y}(\underline{\theta}(\underline{s}), \underline{s}), \underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}^*(\underline{s}), \underline{s}) - C_t(\underline{y}(\underline{\theta}^*(\underline{s}), \underline{s}), \underline{s})] \right] \{\underline{s} | \mathcal{E}\}$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Hence the theorem is proved.

The discussion of Theorem I in Chapter II is also relevant here. Like Theorem I, this theorem under uncertainty only provides sufficient conditions for an optimal solution. However, the theorem requires no restrictive assumptions other than real-valuedness of the functions and it is applicable to problems under very complex forms of uncertainty. Insight into Theorem IV is developed below and in the following sub-sections.

The terms $\lambda_{tk\tau}(\underline{s})$ can be interpreted as prices that depend on the vector of uncertain state variables $\underline{s}$. The price $\lambda_{tk\tau}(\underline{s})$ is the price assigned to the k^{th} resource consumed in period t and paid for in period τ. The additional subscripts on the prices result from simultaneous treatment of multiple resources and time in its most general form. The additional subscripts are <u>not</u> due to the introduction of uncertainty.

In the application of the results of this section, an equivalent set of prices with only two subscripts can be defined. Let,

$$\lambda_{\tau k}(\underline{s}) \equiv \sum_{t=0}^{T} \gamma_t \lambda_{tk\tau}(\underline{s}) .$$

By rearranging the statement of Theorem IV slightly, we can work with prices having only two subscripts. For example, the first maximization problem in Theorem IV becomes:

$$\underset{\underline{\theta}(\underline{s}) \,\in\, \Theta(\underline{s})}{\text{maximize}}$$

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t R_t(\underline{\theta}(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{\tau k}(\underline{s}) y_{\tau k}(\underline{s}) \right] \{\underline{s} \mid \mathcal{E}\} \ .$$

In this form the prices are not multiplied by the discount factors. The price $\lambda_{\tau k}(\underline{s})$ is interpreted as the price (in present value units) assigned to the k^{th} resource consumed or produced in period τ. The choice of which form of the prices to use rests on computational considerations such as the number of storage locations required by the computer program.

The expected present value of profit can be bounded at any stage of an iterative search algorithm. The upper and lower bounds for problems under uncertainty are analogous to the bounds for deterministic problems. The upper bound follows directly from the inequality in Step b) in the proof of Theorem IV.

BOUNDS:

Let $\underline{\theta}'(\underline{s})$ maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s})] \right] \{\underline{s} \mid \mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$ and let $\underline{y}'(\underline{s})$ maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [\sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) y_{\tau k}(\underline{s}) - C_t(\underline{y}(\underline{s}),\underline{s})] \right] \{\underline{s} \mid \mathcal{E}\}$$

over all $\underline{y}(\underline{s})$. Then,

$$\bar{p}^{\ell} = \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\underline{\theta}'(\underline{s}),\underline{s}) - C_t(\underline{y}(\underline{\theta}'(\underline{s}),\underline{s}),\underline{s})] \right] \{\underline{s}|\mathcal{E}\}$$

$$\bar{p}^{u} = \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t [R_t(\theta'(\underline{s}),\underline{s}) - C_t(\underline{y}'(\underline{s}),\underline{s})] \right] \{s|\mathcal{E}\}$$

$$+ \int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[\sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) [y'_{\tau k}(\underline{s}) - y_{\tau k}(\underline{\theta}'(\underline{s}),\underline{s})] \right] \right] \{\underline{s}|\mathcal{E}\} .$$

The successive approximations algorithm provides an interpretation of the prices $\lambda_{tk\tau}(\underline{s})$.

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial 3-dimensional matrix of resource price functions $\underline{\lambda}^0(\underline{s})$, or start at Step 3 with a trial resource policy $\underline{y}^0(\underline{s})$.
2. Maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[R_t(\underline{\theta}(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}^{n}(\underline{s}) y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s}) \right] \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Call the result $\underline{\theta}^n(\underline{s})$.

3. Calculate a new 3-dimensional matrix of resource price functions according to the relationships

$$\lambda_{tk\tau}^{n+1}(\underline{s}) = \frac{\partial}{\partial y_{\tau k}} C_t(\underline{y}(\underline{s}),\underline{s}) \Big|_{\underline{y}(s)=\underline{y}(\underline{\theta}^n(\underline{s}),\underline{s})} \qquad \begin{array}{l} t = 0, \ldots , T , \\ k = 1, \ldots , K , \\ \tau = 0, \ldots , T . \end{array}$$

(Note: $C_t(\ )$ must be convex and differentiable for this algorithm.)

4. If $\underline{\lambda}^{n+1}(\underline{s}) = \underline{\lambda}^n(\underline{s})$ for all logically possible values of $\underline{s}$, then the conditions of Theorem IV are satisfied and $\underline{\theta}^n(\underline{s})$ is the optimal policy. Otherwise, return to Step 2 using $\underline{\lambda}^{n+1}(\underline{s})$.

Step 3 of this algorithm implies that the price function $\lambda_{tk\tau}(\underline{s})$ is the marginal cost in the t^{th} period of the k^{th} resource flow in the τ^{th} period as a function of the state variable vector $\underline{s}$. The function $\underline{\lambda}(\underline{s})$ can be viewed as defining a many-to-many change of variables from $\{\underline{s}|\mathcal{E}\}$ to $\{\underline{\lambda}|\mathcal{E}\}$, the joint probability distribution on the prices. A similar change of variables is defined for the conditional joint probability distributions at any point in time. However, the stochastic process on prices implied by these distributions is usually dependent on the stochastic process on the state variables.

A relaxation version of the algorithm is easily developed. In this case, the price functions are calculated in Step 3 according to the relationship

$$\lambda^{n+1}_{tk\tau}(\underline{s}) = \alpha \frac{\partial}{\partial y_{\tau k}} C_t(\underline{y}(\underline{s}),\underline{s})\Big|_{\underline{y}(\underline{s})=\underline{y}(\underline{\theta}^n(\underline{s}),\underline{s})} + (1-\alpha)\lambda^{n}_{tk\tau}(\underline{s}) \ .$$

The price directive gradient algorithm, with minor modifications, also applies to problems under uncertainty.

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial 3-dimensional matrix of resource price functions $\underline{\lambda}^0(\underline{s})$.

2. Maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[R_t(\underline{\theta}(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda^{n}_{tk\tau}(\underline{s}) y_{\tau k}(\underline{\theta}(\underline{s}),\underline{s}) \right] \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Call the result $\underline{\theta}^n(\underline{s})$.

3. Maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[\sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}^{n}(\underline{s}) y_{\tau k}(\underline{s}) - C_t(\underline{y}(\underline{s}),\underline{s}) \right] \right] \{\underline{s} | \mathcal{E}\}$$

over all $\underline{y}(\underline{s})$. Call the result $\underline{y}^n(\underline{s})$.

4. If $\underline{y}^n(\underline{s}) = \underline{y}(\underline{\theta}^n(\underline{s}),\underline{s})$ for all logically possible values of $\underline{s}$, then the conditions of Theorem IV are satisfied and $\underline{\theta}^n(\underline{s})$ is the optimal policy. Otherwise, compute a new matrix of resource price functions according to

$$\lambda_{tk\tau}^{n+1}(\underline{s}) = \lambda_{tk\tau}^{n}(\underline{s}) - \alpha(y_{\tau k}^{n}(\underline{s}) - y_{\tau k}(\underline{\theta}^n(\underline{s}),\underline{s}))$$

and return to Step 2.

Decomposition

The structural requirements for decomposition under uncertainty are similar to the requirements under certainty. If

$$R_t(\underline{\theta}(\underline{s}),\underline{s}) = \sum_{j=1}^{J} r_{tj}(\underline{\theta}_j(\underline{s}),\underline{s}) \qquad t = 0, \ldots, T,$$

$$x_{tjk}(\underline{\theta}(\underline{s}),\underline{s}) = x_{tjk}(\underline{\theta}_j(\underline{s}),\underline{s}) \qquad \begin{array}{l} t = 0, \ldots, T, \\ j = 1, \ldots, J, \\ k = 1, \ldots, K, \end{array}$$

and

$$\Theta(\underline{s}) = \Theta_1(\underline{s}) \times \cdots \times \Theta_J(\underline{s})$$

where $\underline{\theta}_j(\underline{s}) \in \Theta_j(\underline{s})$ but $\underline{\theta}_j(\underline{s}) \notin \Theta_i(\underline{s})$ for $i \neq j$, then decomposition can be achieved. In this case, Step 2 of the algorithms becomes J independent subproblems, requiring maximization of

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[r_{tj}(\underline{\theta}_j(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) x_{\tau jk}(\underline{\theta}_j(\underline{s}),\underline{s}) \right] \right] \{\underline{s} | \mathcal{E}\}$$

over all $\underline{\theta}_j(\underline{s}) \in \Theta_j(\underline{s})$.

The above result shows that by defining prices it is possible to decompose a hard closed-loop problem under uncertainty into several easier closed-loop problems under uncertainty.† These subproblems can be solved by a variety of techniques including dynamic programming and decision trees.

Each subproblem requires knowledge of the resource price functions and the joint probability distribution on $\underline{s}$. Even if a particular project is deterministic, i.e.,

$$r_{tj}(\underline{\theta}_j(\underline{s}),\underline{s}) = r_{tj}(\underline{\theta}_j(\underline{s})) \qquad t = 0, \ldots, T$$

and

$$x_{tjk}(\underline{\theta}_j(\underline{s}),\underline{s}) = x_{tjk}(\underline{\theta}_j(\underline{s})) \qquad t = 0, \ldots, T, \quad k = 1, \ldots, K$$

the policy for the project usually is stochastic because the resource prices, $\lambda_{th\tau}(\underline{s})$, depend on the uncertain state variables, $\underline{s}$.

Organizational Interpretation

Decomposition under uncertainty can be interpreted in terms of a decentralized organization composed of an impresario, project managers, and resource managers as defined in Section 2.1.

The project managers are responsible for maximizing the expected present value of profit of their projects. The resources required by the project managers are obtained at prices that are given by a stochastic process on the uncertain state variables associated with a project.

† Often it is possible to arrange things so that the subproblems are open-loop problems. Open-loop decision problems are typically much easier to analyze than closed-loop decision problems because the decisions do not depend on the resolution of the uncertainty. When the subproblems are open-loop, the computational advantages of decomposition are considerable.

In the successive approximations algorithm, the resource manager simply computes the marginal cost of the resource as a function of $\underline{s}$. The marginal cost is computed on the basis of the total resource requirements of the projects as a function of $\underline{s}$. If $\underline{s}$ is discrete, the computation of prices is made for each value of $\underline{s}$.

Under uncertainty it is possible to hypothesize additional managers in the decentralized organization. For example, consider the role of the information manager or expert who is responsible for assigning the joint probability distribution, $\{\underline{s} \mid \mathcal{E}\}$. Among his alternatives is the purchase of additional information by experimentation and research. The calculation of the value of information in a decentralized organization or a decomposed problem is an interesting area for future research. One of the objectives of such research might be to carefully define the role of the information manager.

4.2 Problems under Uncertainty with Arbitrary Objective Functions

In this section we demonstrate that problems under uncertainty with arbitrary objective functions can be treated by the methods developed in this dissertation. We will assume that arbitrary objective functions in problems under uncertainty are the result of multi-attribute risk preference functions. A multi-attribute risk preference function can encode both the decision maker's attitude towards uncertain outcomes and his deterministic preferences among multiple measures of performance.

The results of this section complete our development of the mathematical fundamentals of decomposition. Together, these results are applicable to the most general resource allocation problem the author can conceive. Yet, conceptually, the theory in this section is no more difficult than the theory for the separable, single resource problems in Section 2.1.

Introduction to Multi-Attribute Risk Preference Functions

A multi-attribute risk preference function is used to encode a decision maker's attitude towards the outcome of a resource allocation problem under uncertainty. In Section 2.3 ordinal value functions were introduced as a means of encoding a decision maker's attitude towards multiple deterministic outcomes. In this section we extend the introduction in Section 2.3 to problems under uncertainty.

One method of encoding a multi-attribute risk preference function is to encode the decision maker's attitude towards deterministic outcomes and then encode his risk preference. The deterministic part of the problem results in a set of indifference curves as described in Section 2.3. If a numerical index is assigned to the indifference curves so that the resulting ordinal value function has intuitive meaning, then we can encode the decision maker's risk preference in terms of the ordinal value function. For example, if the ordinal value function is in monetary units we can encode the decision maker's risk preference on money. If the ordinal value function is an equivalent uniform flow of a resource, then we can encode the risk preference function by comparing lotteries (probability distributions) on equivalent uniform flow.

This two-step approach to the encoding of multi-attribute risk preference functions is discussed formally in Boyd [5], Boyd and

Matheson [8], Pollard [24], and Raiffa [25].

In this section we use the notation

$$u(\underline{z})$$

to denote a multi-attribute risk preference function defined on $\underline{z}$, the vector of primary resources. The function $u(\underline{z})$ is sometimes written as

$$u(V(\underline{z}))$$

to emphasize the intermediate step of encoding an ordinal value function.

The multi-attribute risk preference function has the property that the optimal resource allocation is the allocation that maximizes the expected value (expected utility)

$$\bar{u} = \int_{\underline{S}} u(\underline{z})\{\underline{z}|\mathcal{E}\}$$

where $\{\underline{z}|\mathcal{E}\} \equiv$ joint probability distribution on the vector of primary resources.

The magnitude of the expected utility resulting from a decision problem has little or no intuitive meaning. Hence, the bounds and prices to be developed in this section would be difficult to interpret in a meaningful way. A measure that does provide insight is the certain equivalent.

A certain equivalent can be developed in terms of the ordinal value function $V(\ )$. Generally, the ordinal value function is designed so that the units of $V(\ )$ have intuitive meaning. If the risk preference function defined on $V(\ )$ has a unique inverse† then the certain equivalent ordinal value is given by

$$\tilde{V}(\bar{u}) = u^{-1}(\bar{u})$$

where $u^{-1}(\bar{u}) \equiv$ inverse of $u(\ )$ such that

$$u(\tilde{V}) = \bar{u} \ .$$

The certain equivalent has the same units as the ordinal value function. For example, an ordinal value function expressed as an equivalent uniform flow corresponds to a certain equivalent uniform flow under uncertainty.

The Example

The notation describing this example builds upon the notation developed in Section 4.1. Let

$\underline{s}_t$ $\equiv$ the vector of uncertain state variables whose uncertainty is resolved at the end of period t.

$\underline{s}$ $\equiv$ the matrix of uncertain state variables where

$$\underline{s} = (\underline{s}_0, \ \ldots \ , \underline{s}_T) \ .$$

$\underline{\theta}(\underline{s})$ $\equiv$ the decision variables describing a policy.

$z_{tk}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ amount of the k^{th} primary resource produced in

---

† $u(\ )$ has a unique inverse if it is monotonic. This condition will almost always be satisfied. In any case, the restriction is imposed only to provide intuitively meaningful bounds and prices. The results of this section can be developed without this monotonicity requirement by working in terms of expected utility rather than certain equivalent.

period t as a function of the policy and the uncertain state variables. This function is given by a detailed structural model of the problem. Often, separate projects can be identified within this structure.

$\underline{z}(\underline{\theta}(\underline{s}),\underline{s})$ $\equiv$ the 3-dimensional matrix of primary resources.

$u(\ )$ $\equiv$ multi-attribute risk preference (utility) function.

$\widetilde{V}(\ )$ $\equiv$ certain equivalent value as a function of the expected utility.

The example problem is to maximize the certain equivalent value

$$\widetilde{V}\left(\int_{\underline{s}} u(\underline{z}(\underline{\theta}(\underline{s}),\underline{s})\{\underline{s}|\mathcal{E}\}\right)$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$ where

$\{\underline{s}|\mathcal{E}\}$ $\equiv$ joint probability density assigned on the basis of the decision maker's prior information $\mathcal{E}$,

and $\Theta(\underline{s})$ $\equiv$ set of all available policies. These policies can be characterized as either delayed, dynamic, or immediate resolution policies.

Mathematical Results (Theorem V, Bounds, and Algorithms)

The following theorem provides sufficient conditions for the optimality of a solution to the example formulated in the previous subsection:

THEOREM V: If $\underline{\theta}^{*}(\underline{s})$ maximizes

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) z_{tk}(\underline{\theta}(\underline{s}),\underline{s}) \right] \{\underline{s} | \mathcal{E}\}$$

over all $\underline{\theta}(s) \in \Theta(\underline{s})$ and if $\underline{w}(\underline{s})$ maximizes

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{w}(\underline{s}))\{\underline{s}|\mathcal{E}\} \right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) w_{tk}(\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{w}(\underline{s})$, and if

$$\underline{w}^*(\underline{s}) = \underline{z}(\underline{\theta}^*(\underline{s}),\underline{s})$$

for all logically possible values of $\underline{s}$, then $\underline{\theta}^*(\underline{s})$ maximizes

$$\tilde{V}\left(\int_{\underline{s}} \left[u(\underline{z}(\underline{\theta}(\underline{s}),\underline{s})) \right] \{\underline{s}|\mathcal{E}\} \right)$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$.[†]

Proof:

a) The conditions of the theorem imply the following two inequalities:

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) z_{tk}(\underline{\theta}(\underline{s}),\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

$$\leq \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) z_{tk}(\underline{\theta}^*(\underline{s}),\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$, and

[†] Theorem V also holds if the risk preference function depends directly on $\underline{s}$, i.e., $u(\underline{z}(\underline{\theta}(\underline{s}),\underline{s}),\underline{s})$. Generally, problems with uncertain values can be reformulated in terms of a risk preference function that does not directly depend on uncertain variables. Pollard [24] provides an introductory discussion of problems with uncertain values.

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{w}(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) w_{tk}(\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

$$\leq \tilde{V}\left(\int_{\underline{s}} u(\underline{w}^*(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) w^*_{tk}(\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

holds for all $\underline{w}(\underline{s})$.

b) Combining the two inequalities in Step a) implies that the inequality

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{w}(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s})[w_{tk}(\underline{s}) - z_{tk}(\theta^*(\underline{s}),\underline{s})] \right] \{\underline{s}|\mathcal{E}\}$$

$$\leq \tilde{V}\left(\int_{\underline{s}} u(\underline{w}^*(\underline{s})\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s})[w^*_{tk}(\underline{s}) - z_{tk}(\underline{\theta}^*(\underline{s}),\underline{s})] \right] \{\underline{s}|\mathcal{E}\}$$

holds for all $w(\underline{s})$.

c) Since the inequality in Step b) holds for all $\underline{w}(\underline{s})$ it must hold for all $\underline{w}(\underline{s}) = \underline{z}(\underline{\theta}(\underline{s}),\underline{s})$ where $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Hence,

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{z}(\underline{\theta}(\underline{s}),\underline{s})\{\underline{s}|\mathcal{E}\}\right)$$

$$\leq \tilde{V}\left(\int_{\underline{s}} u(\underline{w}^*(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s})[w^*_{tk}(\underline{s}) - z_{tk}(\underline{\theta}^*(\underline{s}),\underline{s})] \right] \{\underline{s}|\mathcal{E}\}$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$.

d) The theorem statement requires that

$$\underline{w}^*(\underline{s}) = \underline{z}(\underline{\theta}^*(\underline{s}),\underline{s})$$

hold for all logically possible values of $\underline{s}$. Thus,

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{z}(\underline{\theta}(\underline{s}),\underline{s}))\{\underline{s}|\mathcal{E}\}\right) \leq \tilde{V}\left(\int_{\underline{s}} u(\underline{z}(\underline{\theta}^*(\underline{s}),\underline{s})\{\underline{s}|\mathcal{E}\}\right)$$

holds for all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Hence, the theorem is proved.

This theorem provides the theoretical basis for transforming a difficult problem with a multi-attribute risk preference function into an expected value decision problem with a separable objective function. The resulting expected value problem can be solved by the decomposition methods developed in Section 4.1.

Upper and lower bounds on the certain equivalent value at each stage in an iterative solution process are given by the inequality in Step c) of the proof of Theorem V.

BOUNDS:

Let $\underline{\theta}'(\underline{s})$ maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) z_{tk}(\underline{\theta}(\underline{s}),\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$, and let $\underline{w}'(\underline{s})$ maximize

$$\widetilde{V}\left(\int_{\underline{s}} u(\underline{w}(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) w_{tk}(\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{w}(\underline{s})$. Then,

$$\widetilde{V}^{\ell} = \widetilde{V}\left(\int_{\underline{s}} u(\underline{z}'(\underline{\theta}'(\underline{s}),\underline{s}))\{\underline{s}|\mathcal{E}\}\right)$$

and

$$\widetilde{V}^{u} = \widetilde{V}\left(\int_{\underline{s}} u(w'(\underline{s}))\{\underline{s}|\mathcal{E}\}\right)$$

$$- \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s})[w'_{tk}(\underline{s}) - z_{tk}(\underline{\theta}'(\underline{s}),s)] \right] \{\underline{s}|\mathcal{E}\} .$$

The bounds are expressed in terms of certain equivalent value. If the problem is stated as one of maximizing utility, then the bounds are difficult to interpret.

The successive approximations algorithm follows from the necessary and sufficient conditions for a solution to the second maximization problem in the theorem statement.

SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess an initial matrix of resource price functions $\underline{\mu}^0(\underline{s})$, or start at Step 3 with a trial resource policy $\underline{w}^0(\underline{s})$.
2. Maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}^{n} z_{tk}(\underline{\theta}(\underline{s}),\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Call the result $\underline{\theta}^n(\underline{s})$.

3. Calculate a new matrix of resource price functions according to the relationship

$$\mu_{tk}^{n+1}(\underline{s}) = \frac{\partial}{\partial \bar{u}} \tilde{V}(\bar{u}) \bigg|_{\bar{u}=\bar{u}^n} \cdot \frac{\partial}{\partial w_{tk}} u(\underline{w}) \bigg|_{\underline{w}=\underline{z}(\underline{\theta}^n(\underline{s}),\underline{s})}$$

where $\bar{u}^n = \int_{\underline{s}} u(\underline{z}(\underline{\theta}^n(\underline{s}),\underline{s}))\{\underline{s}|\mathcal{E}\}$.

(Note: $\tilde{V}(u(\underline{z}))$ must be concave and differentiable in $\underline{z}$.)

4. If $\underline{\mu}^{n+1}(\underline{s}) = \underline{\mu}^n(\underline{s})$ for all logically possible values of $\underline{s}$, then, the conditions of Theorem V are satisfied and $\underline{\theta}^n(\underline{s})$ is the optimal policy. Otherwise, return to Step 2.

As in Section 4.1, the operation in Step 3 can be viewed as a many-to-many change of variables problem from $\{\underline{s}|\mathcal{E}\}$ to $\{\underline{\mu}|\mathcal{E}\}$, the joint probability distribution on the prices. These distributions, however, are dependent.

The calculation of the prices in Step 3 involves two terms. The first term is independent of both $\underline{s}$ and the indices t and k. The second term can be interpreted as the marginal utility of the k^{th} resource flow in the t^{th} period as a function of the state variables $\underline{s}$. The first term converts the prices from a marginal utility to a marginal certain equivalent.

In the relaxation version of the successive approximations algorithm the prices are determined by the relationship

$$\mu_{tk}^{n+1}(\underline{s}) = \alpha \frac{\partial}{\partial \bar{u}} \tilde{V}(\bar{u}) \Big|_{\bar{u}=\bar{u}^n} \cdot \frac{\partial}{\partial w_{tk}} u(\underline{w}) \Big|_{\underline{w}=\underline{z}(\theta^n(\underline{s}),\underline{s})} + (1-\alpha)\mu_{tk}^{n-1}(\underline{s})$$

where α is the relaxation coefficient.

The successive approximations algorithm requires that $\tilde{V}(u(\underline{z}))$ be concave and differentiable. Most carefully structured problems meet this requirement. Thus, there appears to be little practical value to the price directive algorithm which does not require the concavity and differentiability restrictions. Nevertheless, for completeness we will state the price directive gradient algorithm for the problem posed in this section.

PRICE DIRECTIVE GRADIENT ALGORITHM:

1. Guess an initial matrix of resource price functions $\underline{\mu}^0(\underline{s})$.
2. Maximize

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}^{n} z_{tk}(\underline{\theta}(\underline{s}),\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{\theta}(\underline{s}) \in \Theta(\underline{s})$. Call the result $\underline{\theta}^n(\underline{s})$.

3. Maximize

$$\tilde{V}\left(\int_{\underline{s}} u(\underline{w}(\underline{s}))\{\underline{s}|\mathcal{E}\}\right) - \int_{\underline{s}} \left[\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_{tk}(\underline{s}) w_{tk}(\underline{s}) \right] \{\underline{s}|\mathcal{E}\}$$

over all $\underline{w}(\underline{s})$. Call the result $\underline{w}^n(\underline{s})$.

4. If $\underline{w}^n(\underline{s}) = \underline{z}(\underline{\theta}^n(\underline{s}),\underline{s})$ for all logically possible values of $\underline{s}$, then the conditions of Theorem IV are satisfied and $\underline{\theta}^n(\underline{s})$ is the optimal policy. Otherwise, compute a new matrix of resource price functions according to

$$\mu_{tk}^{n+1}(\underline{s}) = \mu_{tk}^{n}(\underline{s}) - \alpha[\underline{w}^n(\underline{s}) - \underline{z}(\underline{\theta}^n(\underline{s}),\underline{s})]$$

and return to Step 2.

The price directive gradient algorithm is derived by applying a gradient search to the problem of minimizing the upper bound on the certain equivalent value.

Decomposition

The results of this section show how to decompose the original problem into (1) a problem of determining the appropriate price functions $\mu_{tk}(\underline{s})$ and (2) the new maximization problem in Step 2 of the algorithms. However, the most significant computational advantages arise when we take advantage of the separable structure of this new problem. In some cases this new problem decomposes directly. If not, we can apply the results of Section 4.1 to decompose the new problem. The prices $\underline{\lambda}(\underline{s})$ on the primary resources can be determined simultaneously in a carefully designed algorithm. Conceptually, we now have the means for transforming an extremely difficult closed-loop decision problem under uncertainty with complex preferences into the iterative solution of a number of simple, open-loop, expected value, decision problems.

Organizational Interpretation

The organizational interpretation in Section 4.1 assigned decision making roles to the entrepreneurs and either price setting or decision making roles to the resource managers. With arbitrary objective functions, the impresario can be viewed as the manager of the primary resources. In terms of the successive approximations algorithm the impresario sets the prices on the primary resources.

The impresario can perform his task in two ways. One approach is to encode a multi-attribute risk preference function by the methods mentioned earlier in this section. An alternative approach is to encode the prices directly without explicitly encoding the risk preference function. Under uncertainty, however, the assessment demands placed on the impresario are great because he must determine the prices for every possible value of the state variables.

4.3 Computational Methods for Decomposition under Uncertainty

The essential difference between decomposition of deterministic problems and decomposition under uncertainty is the problem of determining the prices as a function of the state variables, i.e., $\underline{\lambda}(\underline{s})$. In problems with complicated probabilistic structures, effective methods must be developed for calculating and characterizing the price functions. In the most complicated problems, some form of approximation is necessary.

In this section we outline two general approaches to designing computational methods for decomposition under uncertainty. The first method is useful in problems where the state variables are discrete or can be approximated by discrete variables. This method is based on the use of a probability tree to characterize both the probability distribution and the prices.

The second method is useful in problems with continuous variables. This method uses a Taylor series approximation to characterize the price functions.

Probability Tree Methods

Any probabilistic process involving only discrete state variables can be visualized in terms of a probability tree. Sometimes processes with continuous variables can be approximated by a probability tree. Probability trees are particularly useful where the uncertainty concerning state variables is resolved gradually over a number of periods in time.

Figure 4.1 illustrates a simple probability tree. A tree[†] is composed of <u>nodes</u> and <u>branches</u>. Numerical values associated with a tree are called <u>attributes</u>. There are two types of attributes: <u>branch attributes</u> and <u>node attributes</u>. An example of a branch attribute is the probability of going from the node on the starting (left) end of a branch to the node at the terminal (right) end of a branch. An example of a node attribute is the numerical value of the state vector $\underline{s}$ assigned to a node.

The structure of a probability tree together with its attributes describe a probabilistic process. When the node attributes are the state vectors $\underline{s}$ and the branch attributes are the appropriate conditional probabilities, then the tree completely describes the probability distribution $\{\underline{s}|\mathcal{E}\}$ where $\mathcal{E}$ is the state of information at the starting node.

[†] Other types of trees include decision trees and value trees. See Rousseau [27].

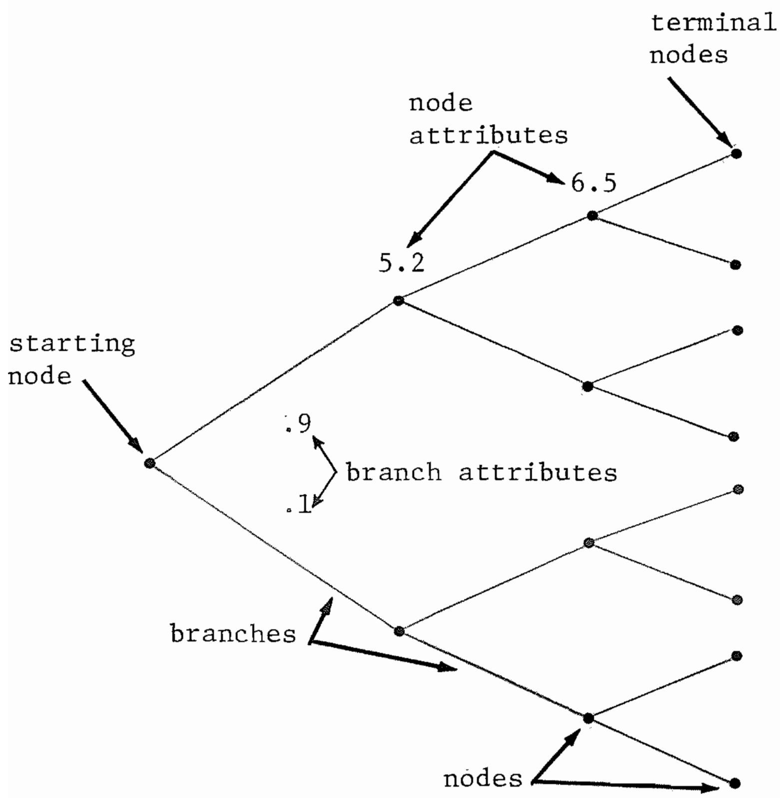


Figure 4.1: A PROBABILITY TREE

Expected values of the variables assigned as attributes are computed by rolling back the tree in an operation similar to dynamic programming.† For example, the operation

$$\int_{\underline{s}} \left[\sum_{t=0}^{T} \gamma_t \left[r_{tj}(\underline{\theta}_j(\underline{s}),\underline{s}) - \sum_{k=1}^{K} \sum_{\tau=0}^{T} \lambda_{tk\tau}(\underline{s}) x_{\tau jk}(\underline{\theta}_j(\underline{s}),\underline{s}) \right] \right] \{\underline{s} \mid \mathcal{E}\}^{\ddagger}$$

can be performed by rolling back a probability tree. For a given policy $\underline{\theta}_j(\underline{s})$, the numerical values of $x_{\tau jk}(\underline{\theta}_j(\underline{s}),\underline{s})$, $r_{tj}(\underline{\theta}_j(\underline{s}),\underline{s})$, and the prices $\lambda_{tk\tau}(\underline{s})$ are simply attributes of the nodes.

The optimization in Step 2 of the algorithms is performed by rolling back a probability tree for each project. In some cases the structure of the problem is such that rolling back a single tree will provide all of the information necessary to determine the expected values for more than one or all of the projects.

In the successive approximations algorithm the prices can be determined by rolling forward through the tree so that every node is reached. As we roll forward we keep track of the resource flows produced by the projects for the current policy. At each node we compute the new prices according to the relationship in Step 3 of the algorithm and assign them as attributes of the appropriate nodes.

Very complicated problems can be solved by carefully structuring the probability tree and the associated computations. Dr. William Rousseau [25] has developed a special compiler that greatly simplifies

---

† Raiffa [25] discusses the roll-back procedure for trees.

‡ This expected value operation is required in the decomposed version of Step 2 of the algorithms in Section 4.1

the task of structuring and manipulating complex trees. An important feature of his compiler is that very large trees with 10^4 or 10^5 nodes can be handled with minimal storage requirements. In some cases, problems under uncertainty require approximately the same amount of computer storage as the equivalent deterministic problem. The principle limitation on the use of probability tree methods in decomposition is the cost of computer processing time.

Approximation Methods

Approximation methods provide an alternative to probability tree methods as a means of characterizing the prices as functions of the uncertain state variables. The basic idea is to approximate the price functions by functions that are easier to characterize.

Considerable ingenuity is often required to devise an approximation to a multi-variable function. Usually, special characteristics of the function must be exploited in order to develop a useful approximation. Thus, the approximations are specific to each problem and it is difficult to provide general methods for approximating the price functions.

In situations where local information provides a good description of a function, approximations based on a Taylor series expansion about an operating point are reasonable. This subsection provides a brief discussion of Taylor series approximations of the price functions. Taylor series methods are unlikely to be useful in the electrical power system example because of the discrete nature of the decision variables. Hence, we will only outline the development of this approximation method.

Specifically, we will develop a Taylor series approximation for the price functions $\underline{\lambda}(\underline{s})$ required in the successive approximations algorithm in Section 4.1.[†] An approximation to the price function using the first three terms of a Taylor series expansion about the operating point $\underline{s}^o$ is given by

$$\lambda_{tk\tau}(\underline{s}) \cong \lambda_{tk\tau}(\underline{s}^o) + \sum_{n=1}^{N} \frac{\partial}{\partial s_n} \lambda_{tk\tau}(\underline{s}) \bigg|_{\underline{s}^o} (s_n - s_n^o)$$

$$+ \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{\partial^2}{\partial s_n \partial s_m} \lambda_{tk\tau}(\underline{s}) \bigg|_{\underline{s}^o} (s_n - s_n^o)(s_m - s_m^o)$$

where $\underline{s} \equiv (s_1, \ldots, s_N)$. This approximation is exact if $\lambda_{tk\tau}(\underline{s})$ is a quadratic or linear function of $\underline{s}$.

When the approximation is valid over the range of $\underline{s}$ encountered in a problem, the coefficients of the Taylor series expansion provide all of the information contained in the price function. Given the coefficients, the optimization problem in Step 2 of the algorithm can be solved. In a computer program, only the coefficients would need to be stored between iterations.

The practical value of Taylor series approximations depends on the structure of the problem and the skill of the analyst in choosing an appropriate set of coefficients. For large numbers of state variables (N large) the number of coefficients is also large. However, the structure of a problem will require that some of the coefficients be fixed at zero. Furthermore, some of the other coefficients will be

---

[†] With appropriate changes in notation the discussion also applies to problems in Section 4.2 with arbitrary objective functions.

small and can be eliminated from the approximation. Sensitivity analysis is useful in deciding where to eliminate coefficients.

By delving more deeply into the structure of a problem we can identify methods for calculating the coefficients of the Taylor series approximation. In the successive approximations algorithm the price functions are given by

$$\lambda_{tk\tau}(\underline{s}) = \frac{\partial}{\partial y_{\tau k}} C_t(\underline{y}(\underline{s}),\underline{s})\Big|_{\underline{y}'(\underline{s})}$$

where $y'_{tk}(\underline{s}) = \sum_{j=1}^{J} x_{tjk}(\underline{\theta}'_j(\underline{s}),\underline{s})$.† This function must be differentiated with respect to the state variables in order to calculate the coefficients. For example, the coefficient

$$\frac{\partial}{\partial s_n} \lambda_{tk\tau}(\underline{s})\Big|_{\underline{s}^o}$$

is given by

$$\frac{\partial^2}{\partial y_{\tau k}\partial s_n} C_t(\underline{y}(\underline{s}),\underline{s})\Big|_{\underline{y}'(\underline{s}^o),\underline{s}^o}$$

$$+ \sum_{u=0}^{T} \sum_{h=1}^{K} \frac{\partial^2}{\partial y_{uh}\partial y_{\tau k}} C_t(\underline{y}(\underline{s}),\underline{s})\Big|_{\underline{y}'(\underline{s}^o),\underline{s}^o} \sum_{j=1}^{J} \frac{\partial}{\partial s_n} x_{ujh}(\underline{\theta}'_j(\underline{s}),\underline{s})\Big|_{\underline{s}^o} .$$

The first term in the above equation is nonzero only when the resource cost function $C_t(\ )$ depends directly on $\underline{s}$. The second term depends on the sensitivity of the consumption of the resources to changes in the state variable s_n. The sensitivity may depend both

---

† We assume that the decomposed version of the algorithm applies and that $\underline{\theta}'_j(\underline{s})$ is the policy for the j^{th} project determined in Step 2 of the algorithm.

on the direct effect of the state variables and on changes in the decision policy $\underline{\theta}_j'(\underline{s})$ as a function of s_n.

The important point to note is that the coefficient can be calculated on the basis of the following data from the projects:

$$x_{\tau jk}(\underline{\theta}_j'(\underline{s}^o),\underline{s}^o) \qquad u = 0, \ldots , T$$

and

$$\frac{\partial}{\partial s_n} x_{ujh}(\underline{\theta}_j'(\underline{s}^o,\underline{s}^o)\Big|_{\underline{s}^o} \qquad h = 1, \ldots , K .$$

The other coefficients require some of the same terms plus some additional second order terms for the second order coefficients. Thus, we see that a Taylor series approximation provides a simple way to characterize the consumption of resources by the projects as a function of the state variables.

Additional computational simplifications are possible, but a detailed discussion of them is beyond the scope of this subsection. One idea is to describe the probability distribution $\{\underline{s}|\mathcal{E}\}$ in terms of the means and covariances of the distribution. For insight as to when this approximation might be useful see Howard [19] on proximal decision analysis. Proximal decision analysis could also be applied in analyzing the subproblems. The sensitivity information required in proximal decision analysis could be used in calculating the coefficients of the Taylor series approximation to the price functions.

CHAPTER V

ELECTRICAL POWER SYSTEM PLANNING UNDER UNCERTAINTY

This chapter has two purposes. Its first purpose is to demonstrate the practicality of decomposition under uncertainty. Power system planning under uncertainty tests the full range of generality provided by the methods of Section 4.1.

The second purpose of this chapter is to evaluate the importance of uncertainty in power system planning. An example is developed that incorporates uncertainty in nuclear fuel prices. The results of the example suggest that other issues are more important than the quantitative analysis of uncertainty in rapidly expanding power systems like the Mexican system.

5.1 The General Problem

The explicit consideration of uncertainty in a power system planning problem is difficult and expensive. Clearly, it is not economic to quantitatively treat every uncertain variable. Many variables have little effect on the installation decisions. Sometimes, the analytical effort is better spent in capturing other aspects of the problem in more detail.

Sensitivity analysis is useful in deciding whether to explicitly treat the uncertainty in a particular variable. In a power system problem, the important decision in an installation policy is the first decision in time; the other decisions serve only as a background policy for the evaluation of the first decision. By varying the state variables in a deterministic model over the range of the uncertainty, we can identify the variables that are critical to the first decision. When we choose to explicitly treat the uncertainty in a variable, then we call it an aleatory variable to distinguish it from variables whose uncertainty is not critical to the first decision.

Several variables in a power system problem are candidates for aleatory variables. Some variables are so obviously important to the decision problem that we include them as aleatory variables at an early stage in the analysis. The reliability model developed in Chapter III provides an example. The available capacity of each plant is an aleatory variable. Similarly, uncertainty in hydro energy and short-term uncertainty in demand can be included in the model developed in Chapter III. Each of these sources of uncertainty are fundamental to power system planning and usually must be included before meaningful results can be obtained.

The remaining candidates for aleatory variables are distinguished by the characteristic that their uncertainty is resolved gradually over time. For example, we expect our forecasts of demand, fuel prices, and capital costs to improve as we approach the period for which the forecast is made.

In a rapidly growing system, small changes in the rate of growth of demand and changes in the trends of other variables can produce spectacular changes in the size and composition of the power system ten years later. When the system growth rate approaches or exceeds the discount rate for a period of time, then the present value index is extremely sensitive to small changes in the growth rate. This suggests that uncertainty in demand growth, for example, is important. However, when uncertainty is resolved gradually over time we are provided the opportunity to adjust installations decisions dynamically in response to the resolution of the uncertainty. In terms of the first decision in an installation policy, the net effect is often small.

In the original analysis of the Mexican system an exhaustive sensitivity analysis was performed on the deterministic model. The sensitivity analysis was closed-loop so that the installation policy could respond to the changes in the state variables. The results of the sensitivity analysis suggested that the explicit consideration of uncertainty would not change the initial decisions in a policy.

In the next section we demonstrate decomposition under uncertainty in terms of uncertainty in nuclear fuel prices. Nuclear fuel price was identified as the most crucial variable in the original analysis primarily because of its effect on the timing of the first installation of a nuclear plant.[†] The results of the example support the suggestion that uncertainty is relatively unimportant in the analysis of installation decisions in the Mexican system.

5.2 A Numerical Example

In this section we introduce uncertainty into the example formulated in Chapter III. The numerical example discussed in this section treats uncertainty in nuclear fuel prices. The example does not provide

[†] Actually, the relative price of nuclear fuel to thermal fuel is the important quantity. For our purposes, we can view forecasts of nuclear fuel prices as conditional forecasts based on known thermal fuel prices.

a complete analysis of uncertainty in the power system problem. The purpose of the example is simply to demonstrate decomposition under uncertainty.

Probabilistic Model of Nuclear Fuel Prices

The first task in introducing an aleatory variable into a model is to develop a probabilistic model of the variable. Probabilistic models of processes evolving over time are difficult to construct. There is no point in developing a more detailed model than is economic in terms of the first decision in a policy. The model developed in this subsection incorporates a level of detail appropriate to a pilot or first-cut analysis of uncertainty in nuclear fuel prices.

Figure 5.1 contains a probability tree describing the model of nuclear fuel prices. Every three years the annual rate of change in the fuel price is subject to change. The probabilities assigned to the branches of the tree reflect the tendency of the rate to remain steady rather than to fluctuate.

The nuclear fuel price in the numerical example in Section 3.5 decreases at a nominal rate of 1.7 per cent per year. In this model, the probabilities are assigned to the branches of the tree so that the expected rate of change is approximately the same. The expected price of fuel in 1985, based on the information available in 1969, is 74 per cent of the 1969 price. The standard deviation of the price in 1985 is 12 per cent of the 1969 price. The data for this model was generated by the author and does not reflect the expertise that would be available from a person familiar with the nuclear fuel markets.

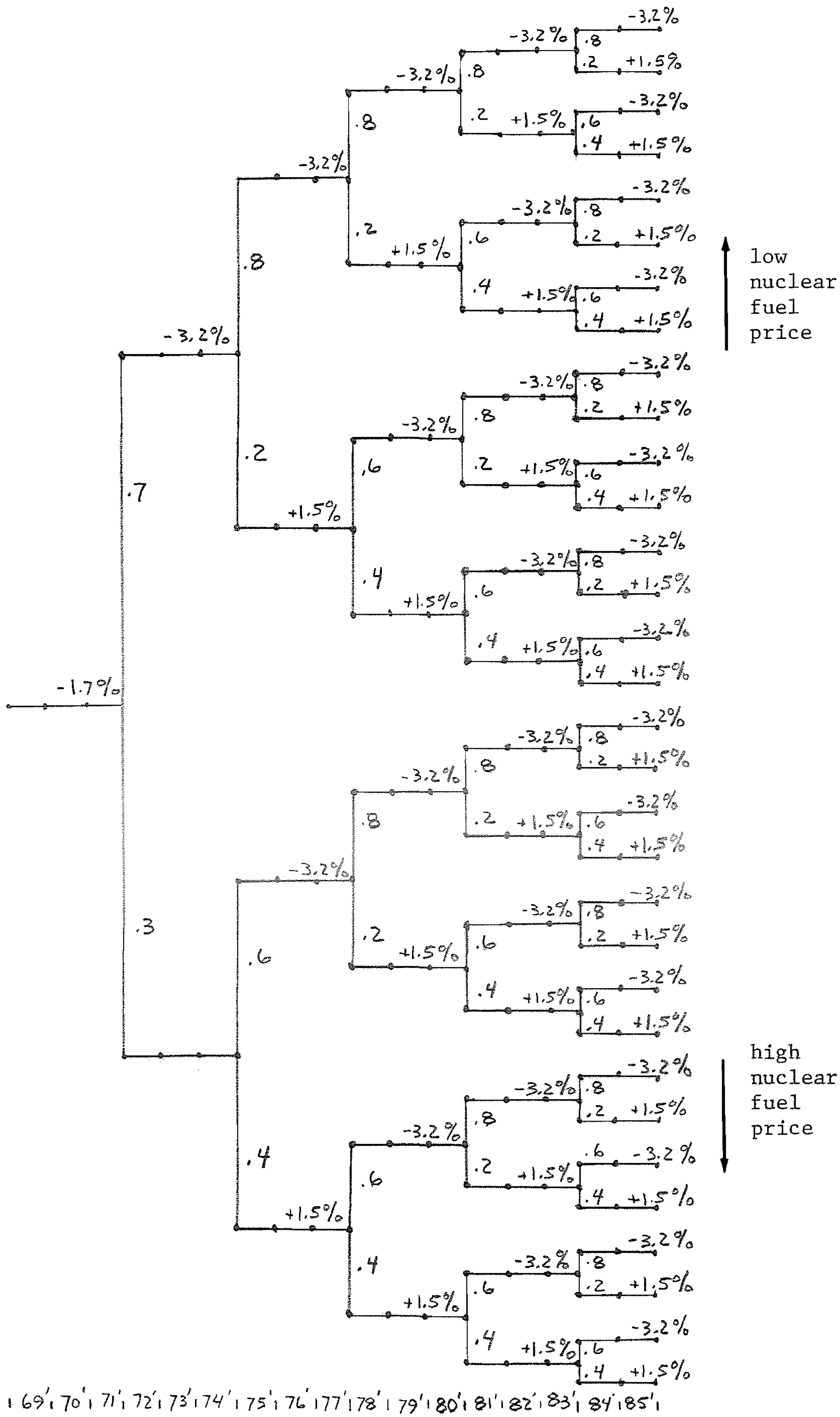


Figure 5.1: PROBABILISTIC MODEL OF NUCLEAR FUEL PRICES

In a full-scale analysis the probabilistic model might be based on a detailed analysis of the nuclear fuel markets. The events that cause changes in prices would be identified and probabilities would be assigned to these events. The interdependence of the nuclear and fossil fuel markets could be modeled in detail. Since the purpose of this example is only to demonstrate decomposition under uncertainty, this hypothetical model is adequate.

Formulation of the Algorithm

The power system problem with uncertainty in nuclear fuel prices can be viewed as a simple extension of the deterministic example developed in Chapter III. Let the vector

$$\underline{s} = (s_0, \ldots , s_T)$$

define the price of nuclear fuel in each period. The probabilistic process that generates the prices is given by the joint probability distribution

$$\{\underline{s}|\mathcal{E}\} .$$

The uncertainty in nuclear fuel prices directly affects the total hourly operating cost of nuclear plants.

By taking advantage of the insights developed in Section 4.1, we can write a decomposition algorithm directly. The sequential successive approximations algorithm with uncertainty in nuclear fuel prices is as follows:

SEQUENTIAL SUCCESSIVE APPROXIMATIONS ALGORITHM:

1. Guess initial prices $\lambda_t^c(\underline{s})$, $\lambda_t^{\bar{c}}(\underline{s})$, $\lambda_t^{\check{c}}(\underline{s})$, $\lambda_{it}^{oc}(\underline{s})$, and $\lambda_{it}^{oh}(\underline{s})$ where $i = 0, \ldots, I$, and $t = 0, \ldots, T$ or start with an initial policy $\underline{\theta}(\underline{s})$ at Step 3.

2. Maximize

$$\int_{\underline{s}} \Bigg[\sum_{t=\tau}^{T} \gamma_t[-f_\tau(t,\theta_\tau(\underline{s})) - y_\tau(t,\theta_\tau(\underline{s}))]$$

$$+ O_\tau\Big(\sum_{t=0}^{\tau} \underline{c}_t(t,\theta_t(\underline{s})),\ \sum_{t=0}^{\tau} \underline{\kappa}_\tau(t,\theta_\tau(\underline{s}),\underline{s})\Big)$$

$$+ C_\tau\Big(\sum_{t=0}^{\tau} c_t(t,\theta_t(\underline{s})),\ \sum_{t=0}^{\tau} \bar{c}_t(t,\theta_t(\underline{s})),\ \sum_{t=0}^{\tau} \check{c}_t(t,\theta_t(\underline{s}))\Big)$$

$$+ \sum_{t=\tau+1}^{T} \gamma_t\Big[\lambda_t^c(\underline{s})c_t(t,\theta_\tau(\underline{s})) + \lambda_t^{\bar{c}}(\underline{s})\bar{c}_t(t,\theta_\tau(\underline{s}))$$

$$+ \lambda_t^{\check{c}}(\underline{s})\check{c}_t(t,\theta_\tau(\underline{s}))$$

$$+ \sum_{i=0}^{T} [\lambda_{it}^{oc}(\underline{s})c_{it}(t,\theta_\tau(\underline{s})) + \lambda_{it}^{oh}\kappa_{i\tau}(t,\theta_\tau(\underline{s}),\underline{s})]\Big]$$

$$+ \gamma_{T+1}v_\tau(T+1,\theta_\tau)\Bigg] \{\underline{s}|\mathcal{E}\}$$

over all $\theta_\tau(\underline{s}) \in \Theta_\tau(\underline{s})$ for $\tau = 0$. Repeat for $\tau = 1, \ldots, T$ in ascending order of the index τ. Call the results $\theta_\tau^n(\underline{s})$.

3. Calculate new prices according to the relations

$$\lambda_t^c(\underline{s}) = -\frac{\partial}{\partial x_t} C_t(x_t,\bar{x}_t,\check{x}_t)\Big|_{x_t^n(\underline{s}),\bar{x}_t^n(\underline{s}),\check{x}_t^n(\underline{s})}$$

$$\lambda_t^{\bar{c}}(\underline{s}) = -\frac{\partial}{\partial \bar{x}_t} C_t(x_t,\bar{x}_t,\check{x}_t)\Big|_{x_t^n(\underline{s}),\bar{x}_t^n(\underline{s}),\check{x}_t^n(\underline{s})}$$

$$\lambda_t^{\check{c}}(\underline{s}) = -\frac{\partial}{\partial x_t} C_t(x_t,\bar{x}_t,\check{x}_t)\Big|_{x_t^n(\underline{s}),\bar{x}_t^n(\underline{s}),\check{x}_t^n(\underline{s})}$$

$$\lambda_{it}^{oc}(\underline{s}) = -\frac{\partial}{\partial x_{it}} O_t(\underline{x}_t,\underline{h}_t)\Big|_{\underline{x}_t^n(\underline{s}),\underline{h}_t^n(\underline{s})}$$

$$\lambda_{it}^{oh}(\underline{s}) = -\frac{\partial}{\partial h_{it}} O_t(\underline{x}_t,\underline{h}_t)\Big|_{\underline{x}_t^n(\underline{s}),\underline{h}_t^n(\underline{s})}$$

where

$$x_t^n(\underline{s}) = \sum_{\tau=0}^{t} c_\tau(t,\theta_\tau^n(\underline{s}))$$

$$\bar{x}_t^n(\underline{s}) = \sum_{\tau=0}^{t} \bar{c}_\tau(t,\theta_\tau^n(\underline{s}))$$

$$\check{x}_t^n(\underline{s}) = \sum_{\tau=0}^{t} \check{c}_\tau(t,\theta_\tau^n(\underline{s}))$$

$$\underline{x}_t^n(\underline{s}) = \sum_{\tau=0}^{t} \underline{c}_\tau(t,\theta_\tau^n(\underline{s}))$$

$$\underline{h}_t^n(\underline{s}) = \sum_{\tau=0}^{t} \underline{\kappa}_\tau(t,\theta_\tau^n(\underline{s}),\underline{s}) .$$

4. If the new prices equal the prices determined on the previous iteration (for all logically possible values of $\underline{s}$), then the optimal policy is $\underline{\theta}^n(\underline{s})$. Otherwise, return to Step 2 using the new prices computed in Step 3.

Implementation of the Algorithm

The algorithm was implemented on the computer by using the probability tree methods discussed in Section 4.3. The program was written so that deterministic runs could be made simply by changing the input data. The computer program is identical to the program developed in

Chapter III, although some features of the program are unnecessary for the deterministic example.

As in the deterministic example, the decision routine evaluates the combinations of plants in the catalog for installation in each year. Under uncertainty, it is important to carefully define the state of information at the time a decision is made. Thus, the lead-time between the decision to install a plant and the first operation of the plant is important under uncertainty.

The lead-time between the decision to install a plant and its first operation depends on the type of plant. Gas turbines typically require two years or less lead-time; thermal units about three years; and nuclear units about five or six years. In the numerical example, we assume all decisions are made with a six-year lead-time.† A uniform lead-time for all plants simplifies the design of the catalog under uncertainty. If a sequential algorithm was not required to overcome gaps, then there would be no need for the uniform lead-time assumption since installation decisions could be made independently for each type of plant. The assumption is justified because the relative insensitivity of the results to the uncertainty in nuclear fuel prices does not justify the cost of a more complicated decision routine for the sequential algorithm.

Results of the Numerical Example

The optimal policy under uncertainty in nuclear fuel prices is

---

† The lead-time is provided as an input parameter to the program and can be changed.

summarized in Figure 5.2. Only the installation decisions are shown in this tree. The optimal policy was achieved in four iterations[†] using a relaxation coefficient of 0.5. The initial policy is the optimal policy from the deterministic example in Chapter III. The difference in the expected present value of profit between the initial policy and the optimal policy in Figure 5.2 is only 1.2 million dollars. The present value of the deterministic example was 1194.2 million dollars.

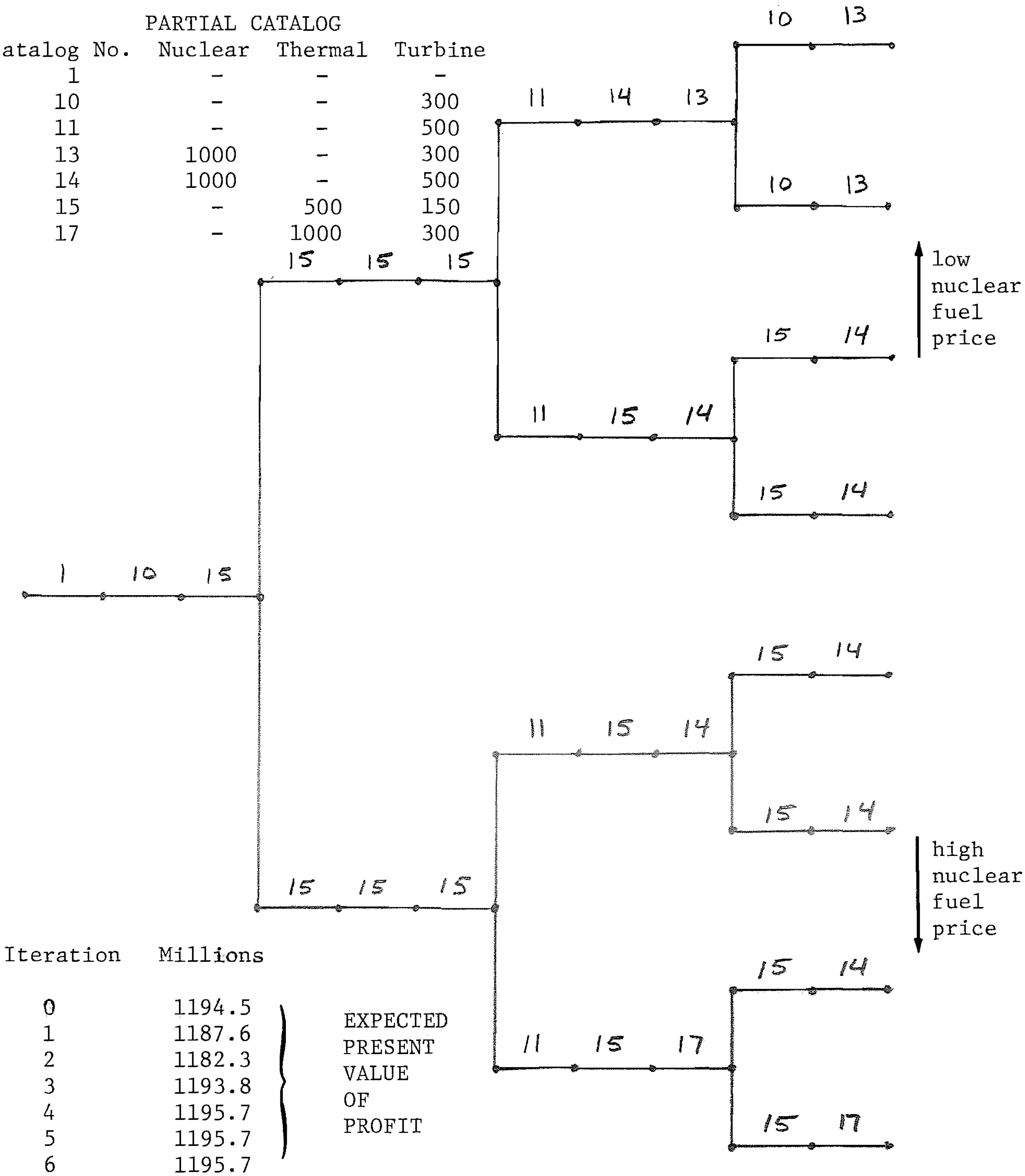
The most important feature of the optimal policy under uncertainty is that the initial decisions are the same as in the deterministic policy. Changes in the policy do not occur until 1982 when an additional nuclear plant is installed in the situation with the lowest nuclear fuel price. The changes in the policy after 1982 are in the directions our intuitions would suggest (more nuclear capacity when nuclear fuel price is low, more thermal capacity when nuclear fuel price is high).

The insensitivity of the initial decisions to uncertainty in nuclear fuel price supports the conclusion that uncertainty is relatively unimportant in planning rapidly expanding power systems. The uncertainty might be more important if nuclear plants were among the initial installations in the optimal deterministic policy. Nuclear plants operate at relatively high load factors. Increases in nuclear fuel prices cannot be offset by installing more efficient plants. Nevertheless, it seems unlikely that the basic conclusion concerning uncertainty would be changed when the initial installations include nuclear plants.

[†] The total cost of the computer run summarized in Figure 5.2 was approximately $160.00. Prices on 1332 resources are calculated on each iteration.

PARTIAL CATALOG

Catalog No.	Nuclear	Thermal	Turbine
1	-	-	-
10	-	-	300
11	-	-	500
13	1000	-	300
14	1000	-	500
15	-	500	150
17	-	1000	300



Iteration	Millions
0	1194.5
1	1187.6
2	1182.3
3	1193.8
4	1195.7
5	1195.7
6	1195.7

EXPECTED PRESENT VALUE OF PROFIT

75'	76'	77'	78'	79'	80'	81'	82'	83'	84'	85'
1	10	15	15	15	15	11	15	14	15	14

Figure 5.2: OPTIMAL POLICY UNDER UNCERTAINTY IN NUCLEAR FUEL PRICES

Conclusions Based on the Example

In this chapter we have explored the effects of uncertainty in power system planning. On the basis of a previous analysis we selected nuclear fuel prices as aleatory variables for the numerical example. The results of the numerical example suggest the following conclusion: Uncertainty that is resolved gradually over time is unimportant for capacity expansion planning in rapidly growing power systems.[†] Thus, additional analytical effort would be more effective if it were spent on extending the scope of the model as suggested in Section 3.7.

[†] Of course, this conclusion depends on the degree of uncertainty, but the general insensitivity of the initial decisions to the uncertainty remains.

CHAPTER VI

SUMMARY AND CONCLUSIONS

6.1 Summary

One of the objectives of this dissertation is to develop a methodology for solving complex strategic decision problems in situations where detailed models are required. The other objective is to apply the methodology to electrical power system planning.

The mathematical foundations of decomposition are developed by using a series of five resource allocation problems. The first problem involves the allocation of a single resource among a number of projects under deterministic conditions. The second problem is a multiple resource problem. If time is modeled in discrete periods, dynamic problems can be viewed as multiple resource problems. The third resource allocation problem has an arbitrary (nonseparable) objective function. Arbitrary objective functions arise in problems with multiple measures of performance. Uncertainty is introduced in the fourth problem. Finally, the fifth problem involves both arbitrary objective functions and uncertainty.

Together, the five resource allocation problems incorporate every aspect of complex decision problems. An important result of this dissertation is that each of the five problems can be decomposed and solved using the same basic techniques. Furthermore, these techniques require no advanced mathematics beyond elementary calculus.

The five related optimality theorems developed in this dissertation provide the theoretical basis for all of the results of this dissertation. The theorems provide a "fail-safe" test for the optimal solution of a resource allocation or decision problem. Any solution satisfying the conditions of the theorem is guaranteed to be a global optimum. The theorems are applicable to problems with discrete and nonlinear functions.

In their simplest form, the optimality theorems are related to a theorem popularized by Everett [12]. Everett's theorem was developed for constrained problems under certainty. In this dissertation we reinterpret Everett's theorem in terms of an unconstrained optimization problem and extend its application to problems with arbitrary objective functions and complex forms of uncertainty.

Two basic algorithms that iteratively search for the optimal solution are suggested by the optimality theorems. The successive approximations algorithm follows directly from the unconstrained interpretation of an optimization problem in terms of projects that consume or produce resources for sale or purchase in a resource market. At each stage in an iterative search, upper and lower bounds can be computed. An important feature of the algorithms is that intermediate, suboptimal solutions are feasible and can be implemented if desired.

Conceptually, the mathematical results of the theorem are valid for very general problems. In some problems, gaps make the results less useful. Several methods for treating gaps are discussed in this dissertation. The solution of the electrical power system problem in the presence of gaps was achieved by modifying the algorithms.

The results on decomposition under uncertainty are applicable to problems with very complex formsof uncertainty. The notation used to describe problems under uncertainty greatly simplifies the analysis. Powerful computational techniques for decomposition under uncertainty are proposed. The literature on mathematical programming and economics has very few nontrivial results on decomposition under uncertainty.

The sections on arbitrary objective functions extend decomposition to problems with multiple measures of performance. Thus, the methodology developed in this dissertation is applicable to problems with complex, nonmonetary objective functions. The decomposition approach also provides insight into the problem of structuring preferences and assessing values.

The results of this dissertation also suggest a method for identifying the resources, resource markets, and projects that permit decomposition. The application of the method to electrical power system planning illustrates that problems with complex technical interactions can be solved by decomposition.

The organizational interpretation of decomposition in terms of decentralized organizations provides many insights. For example, externalities in an economy are analogous to the complex technical interactions in a power system problem. Insight into decision making in organizations where profit is not the sole consideration is provided by the sections on arbitrary objective functions. The results of this dissertation can be viewed as contributing to the mathematical theory of decentralized organizations. However, none of the practical aspects of decentralization were explored.

The electrical power system example demonstrates that every aspect of this complicated strategic decision problem can be treated in a practical way. Decomposition provides insight into power system planning and suggests directions for extending the scope of the model. The solution of the power system problem under uncertainty demonstrates the practicality of decomposition under uncertainty. The results of the example suggest the following conclusion: Uncertainty that is resolved gradually over time is unimportant to capacity expansion planning in rapidly growing power systems.

6.2 Directions for Future Research

The results of this dissertation open up several new areas for future research. The following is a partial list:

1. Additional algorithms are clearly possible. Algorithms that take full advantage of previous results in generating new solutions would be valuable.
2. Further theoretical study of the convergence of decomposition algorithms would provide insight for choosing the best algorithm for a given problem.
3. Tactics for applying penalty function methods in the most effective way would be useful in solving problems with gaps.
4. Further development of approximation methods for decomposition of problems under uncertainty appears possible.
5. The entire subject of information value theory in the context of decomposed problems and decentralized organizations has not been investigated.
6. A general computer program could be written that would solve

general nonlinear and discrete optimization problems under uncertainty. The inputs to the program would describe the resources, projects, and resource markets associated with an optimization problem.

7. The results of this dissertation provide insight into resource allocation methods for decentralized corporations. For example, a capital budgeting system could be designed to operate on an iterative basis using prices. Research into the practical aspects of such systems would be valuable.

8. The design of new institutions in a society can be viewed as an application of the results of this dissertation.† Research into the practical aspects of this problem would be valuable.

6.3 Conclusions

This dissertation is an application of the "divide-and-conquer" philosophy of problem solving: When faced with a complex problem, break it down into small parts which can be understood and analyzed and then put the parts back together. We have demonstrated this philosophy in our research on decomposition. The essential theoretical results on decomposition were developed in terms of simple examples and then extended in small steps to problems under uncertainty. The results on decomposition extend the divide-and-conquer philosophy to the optimization of complex decision problems. The ultimate application of this philosophy will be its use in the design of new institutions for society.

† For example, see the brief discussion at the end of Section 3.7. Some additional insights into this problem are contained in Boyd and Cazalet [6].

REFERENCES

1. Arrow, Kenneth J., "Political and Economic Evaluation of Social Effects and Externalities," Unpublished Paper, Stanford University, 1969.

2. Arrow, Kenneth J. and L. Hurwicz, "Decentralization and Computation in Resource Allocation," in <u>Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling</u>, R. W. Pfouts ed., Chapel Hill: Univ. of North Carolina Press, 1960, pp. 34-104.

3. Bellmore, M., H. J. Greenberg, and J. J. Jarvis, "Generalized Penalty Function Concepts in Mathematical Optimization," <u>Operations Research</u>, Vol. 18, No. 2, March-April 1970, pp. 229-252.

4. Berrie, T. W., "The Economics of System Planning in Bulk Electricity Supply," <u>Electrical Review</u>, Vol. 181, 15th, 22nd, and 29th Sept., 1967, reprinted in <u>Public Enterprise</u>, R. Turvey ed., Penguin Books, Middlesex, England, 1968.

5. Boyd, D. W., "A Methodology for Analyzing Decision Problems Involving Complex Preference Assessments," Ph.D. Dissertation, Engineering-Economic Systems Department, Stanford University, May 1970.

6. Boyd, D. W. and E. G. Cazalet, "Resource Allocation Using Price Iteration for Decentralization: Basic Concepts," Stanford Research Institute Internal Research Report, draft completed

November 1969, final version in preparation will incorporate the results in this dissertation and in the dissertation of Boyd [5].

7. Boyd, D. W. and E. G. Cazalet, "Resource Allocation Using Price Iteration for Decentralization: Summary," Proceedings of Third Hawaii International Conference on System Sciences, January 1970.

8. Boyd, D. W. and J. E. Matheson, "Multiproject Selection Decisions: Problem Definition," Stanford Research Institute Internal Research Report, January 1969.

9. Cazalet, Edward G., Unpublished project memos concerning the reliability of the CFE system, Stanford Research Institute, 1968.

10. Cazalet, E. G. et al, "Decision Analysis of Nuclear Plants in Electrical System Expansion," Final Stanford Research Institute Report to Comisión Federal de Electricidad, Mexico City, D. F., December 1968.

11. Courant, R., Differential and Integral Calculus, Blackie, London, 1934.

12. Everett, H., III, "Generalized Lagrange Multiplier Method for Solving Problems of Optimal Allocation of Resources," Operations Research, Vol. 11, No. 3, May-June 1963, pp. 101-111.

13. Fernandez, G. and J. Eibenschutz, "Modeling Philosophy in Preparation of a Global Planning Model for a Government Utility," Presented at Third Hawaii International Conference on System Sciences, January 1970.

14. Fiacco, A. V. and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, Inc., New York, 1968.

15. Geoffrion, A. M., "Primal Resource-Directive Approach for Optimizing Nonlinear Decomposible Systems," Rand Corporation Memorandum RM-5829-PR, December 1968.

16. Gnedenko, B. V., The Theory of Probability, translated from the Russian by B. D. Seckler, Chelsea Publishing Company, New York, 1962.

17. Helms, Eugene, "Investment for the Firm as a Resource Management Problem," Ph.D. Dissertation, Engineering-Economic Systems Department, Stanford University, in preparation.

18. Howard, R. A., "The Foundations of Decision Analysis," IEEE Transactions on Systems Sciences and Cybernetics, Vol. SSC-4, No. 3, September 1968, pp. 211-219.

19. Howard, Ronald A., "Proximal Decision Analysis," Unpublished paper, Stanford University, April 1970.

20. Lasdon, L. S., "Duality and Decomposition in Mathematical Programming," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 2, July 1968, pp. 86-99.

21. Mitchell, David L., "Two-Stage Allocation Problems with Uncertain Second-Stage Resources - A Nonlinear Programming Problem," Ph.D. Dissertation, Engineering-Economic Systems Department, Stanford University, March 1968.

22. Morris, William T., Decentralization in Management Systems: An Introduction to Design, Ohio University Press, 1968.

23. Nelson, J. R. ed., "Marginal Cost Pricing in Practice," Prentice-Hall, Inc., 1964.

24. North, D. W., "A Tutorial Introduction to Decision Theory," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, September 1968, pp. 200-210.

25. Pollard, A. B., "A Normative Model for Joint Time/Risk Preference Decision Problems," Ph.D. Dissertation, Engineering-Economic Systems Department, Stanford University, August 1969.

26. Raiffa, Howard, Decision Analysis: Introductory Lectures on Choices under Uncertainty, Addison-Wesley, Reading, Massachusetts, 1968.

27. Rousseau, William R., "Tree Generation and Analysis," Stanford Research Institute Internal Research Report, 1968.

28. Turvey, Ralph, Optimal Pricing and Investment in Electricity Supply, MIT Press, Cambridge, Massachusetts, 1968.

29. Wilson, Robert B., "Decomposition Methods in Dynamic Programming," Working Paper No. 108, Graduate School of Business, Stanford University, July 1966.