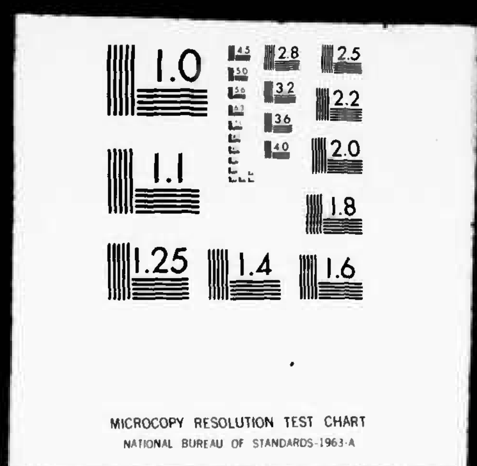
AD-A058 421 MASSACHUSETTS INST OF TECH CAMBRIDGE OPERATIONS RESE--ETC F/G 12/2
DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY --ETC(U)
AUG 78 J A BLOOM DAAG29-76-C-0064
UNCLASSIFIED TR-154 ARO-14261.10-M NL

1 OF 4
AD
A058421

1 OF 4

AD
A058421



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ~~Technical Report No. 154~~	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER (9)
4. TITLE (and Subtitle) (6) DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY PLANNING MODELS		5. TYPE OF REPORT & PERIOD COVERED Technical Report, August 1978
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Jeremy A. Bloom		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-76-C-0064, E(11-1)-4119
9. PERFORMING ORGANIZATION NAME AND ADDRESS M.I.T. Operations Research Center 77 Massachusetts Avenue Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ~~P-14261-M~~
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE (11) August 1978
		13. NUMBER OF PAGES 317 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (14) TR-154 (12) 320p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Releasable without limitation on dissemination.

(18) ARO

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(19) 14261.10-M

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Mathematical Programming Decomposition
Capacity Expansion
Probabilistic Simulation
Peakload Pricing
Nondifferentiable Optimization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See page 2.

78 09 01 022

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

270 720

DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY PLANNING MODELS

by

JEREMY A. BLOOM

Technical Report No. 154

Work Performed Under

Contract DAAG29-76-C-0064, U.S. Army Research Office

"Basic Studies in Combinatorial and Nondifferentiable Optimization"

M.I.T. OSP 84475

Operations Research Center

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

August 1978

FOREWORD

The Operations Research Center at the Massachusetts Institute of Technology is an interdepartmental activity devoted to graduate education and research in the field of operations research. The work of the Center is supported by government grants and contracts. The work reported, herein, was supported, in part, by the U.S. Army Research Office under Contract DAAG29-76-C-0064.

The author also wants to acknowledge research support from the M.I.T. Energy Laboratory through Department of Energy Contracts E(11-1)-4119 and EX-76-A-01-2295.

ABSTRACT

This paper presents several mathematical programming models for planning expansion of electricity generating capacity by utilities. The objective considered is to minimize the cost of meeting a given set of demands over a multi-period planning horizon. In this formulation, the problem naturally decomposes into two parts - determining the optimal plant capacity investments over the entire horizon and determining the optimal operating schedule for the generating plants in each period.

This paper discusses how mathematical programming decomposition techniques can be used to exploit this natural decomposition. Because the operating problems often have simple structure which can be solved essentially in closed form, efficient decomposition algorithms for the entire problem can be formulated. Three related models are presented - one based on linear programming, one based on nonlinear programming for the case when plants are completely reliable, and one based on nonlinear programming for the case when plants can fail randomly. In this probabilistic case, the technique of probabilistic simulation is used to determine expected operating costs and system reliability.

The paper also discusses how these models can be used in peak-load pricing. The probabilistic capacity planning model can be used to calculate the marginal costs attributable to demand at different times. These marginal costs can be used in an equilibrium problem to determine peak-load prices. The equilibrium problem can be solved by a decomposition approach in which the capacity planning model is used as a subproblem.

Acknowledgements

I would like to express my gratitude for the guidance and support I received while writing this thesis to

Professor Jeremy Shapiro, my thesis adviser, for his encouragement and insight at crucial points in my research;

Professors Henry Jacoby, Fred Schweppe, and Dan Kleitman, who served on my thesis committee, for their interest in my work and their useful suggestions;

Ed Moriarty, for his extraordinary efforts in helping me with the computation for my thesis, my sincerest thanks;

Susan Finger, who originally suggested the topic for my thesis and who provided many useful ideas along the way;

Roberta Lorn, Mark Snyder, Mitchell Wangh, and my good friends and colleagues at the Operations Research Center for their sympathy, interest, and encouragement of my work;

Phyllis Ruby, for typing my thesis.

This thesis is lovingly dedicated to my parents Harold and Rosalie Bloom.

Table of Contents

Chapter 1 Introduction and Summary 9

Part One Mathematical Programming Models for Electric Utility Capacity Planning 29

Introduction

A. Organization of Part One 30

B. Definitions and Problem Statement 33

Chapter 2 A Linear Programming Model 40

A. Introduction 40

B. Optimal Operation of the Generating System 43

C. An Optimal Generating Capacity Expansion Model and Its Solution by Benders' Decomposition 50

Chapter 3 A Nonlinear Programming Model for the Deterministic Case 59

A. Introduction 59

B. The Deterministic Operating Problem 62

C. The Capacity Expansion Planning Problem and Its Solution by Generalized Benders' Decomposition 70

Chapter 4 A Model for the Probabilistic Case 76

A. Introduction 76

B. Probabilistic Simulation and the Operating Problem 79

C. The Capacity Expansion Planning Model and Its Solution by Generalized Benders' Decomposition 93
D. An Alternative Treatment of Reliability 99
Chapter 5 Some Computational Results 102
A. Introduction 102
B. Implementation 103
C. Results of Test Runs 107
Chapter 6 Extensions 120
A. Introduction 120
B. Hydroelectric Plants 121
C. Nonlinear Capacity Costs and Fixed-Size Plants 126
D. Multi-Valve-Point Plants and Spinning Reserves 129
E. Maintenance Scheduling 131
F. Demand Uncertainty 132
Part Two Technical Issues 134
Introduction
A. Organization of Part Two 135
B. Results from Duality Theory 138
Chapter 7 Some Technical Issues Associated with the Deterministic Model 142
A. Derivation of the Master Problem 142

	B. Solution of the Subproblem	152
	C. Subproblem Infeasibility	157
	D. Subproblem Degeneracy	160
Chapter 8	Some Technical Issues Associated with the Probabilistic Model	162
	A. Derivation of the Master Problem	162
	B. Solution of the Subproblem	171
	C. Subproblem Infeasibility	185
	D. Subproblem Degeneracy	189
Chapter 9	Computational Methods for the Probabilistic Operating Subproblem	191
	A. Introduction	191
	B. Determining the Marginal Plant and Computing Expected Operating Cost and Energy	196
	C. Computation of the Shadow Prices	199
	D. Outline of the Complete Algorithm	210
	<u>Part Three Application of the Decomposition Approach to Peak-Load Pricing</u>	213
Introduction		214
Chapter 10	Marginal Cost Calculations Using the Capacity Planning Model	220
	A. Introduction	220
	B. Hour-by-Hour Allocation of Loss-of-Load Probability	224

C. Determination of Hourly Contributions to Cost and Unserved Energy 230

D. A Model for Determining Supply Cost and Capacity for Planning 238

Chapter 11 Some Technical Issues 246

A. Introduction 246

B. Some General Results 249

C. Application to the Capacity Planning Problem 260

Chpater 12 Calculating Peak-Load Prices 276

A. Introduction 276

B. An Equilibrium Model for Peak-Load Pricing 281

C. Some Practical Issues 289

Footnotes 295

Bibliography 305

Index of Notation 311

List of Figures

Figure 3.1	Derivation of the Load Duration Curve	63
Figure 3.2	Merit Order Operation	65
Figure 4.1	A Typical History of Plant Operation	80
Figure 4.2	Derivation of the Equivalent Load Duration Curve	83
Figure 4.3	Loss-of-Load Probability and Expected Unserved Energy	86
Figure 5.1	The Algorithm for Solving the Probabilistic Capacity Planning Problem	105
Table 5.1	Problem Summary	108
Table 5.2	Plant Data	109
Figure 5.2	Convergence of the Algorithm in Problem A	111
Table 5.3	Selected Trial Plans from Problem A	112
Figure 5.3	Convergence of the Algorithm in Problem B	113
Table 5.4	Selected Trial Plans from Problem B	114
Figure 5.4	Convergence of the Algorithm in Problem C	115
Table 5.5	Selected Trial Plans in Problem C	116
Figure 6.1	Loading a Hydroelectric Plant	124
Figure 10.1	Derivation of the Probability Function g_i	226
Figure 10.2	Derivation of the Probability Function h_i	228

CHAPTER 1

INTRODUCTION AND SUMMARY

Like other types of industrial firms, electric utilities have the problem of planning their capital investments in generating (or manufacturing) plant to meet the future demands of their customers. In general terms, the problem to be considered is to find a minimum cost capacity expansion plan to meet forecast demands for electricity over a multi-year planning horizon, usually twenty to thirty years. Cost in this problem has two components, the initial capital cost of each plant to be built and the continuing cost of operating the system of generating plants to meet the customers' demand.

An important characteristic of the demand for electricity is that it is highly time-dependent, varying considerably both in the course of a day and in different seasons of the year. Thus, though enough generating capacity must be built to meet the peak levels of demand, some of that capacity will be idle during a significant fraction of the time. Therefore, minimizing the cost of meeting demand involves trading-off the capital costs of the plants, which for a given type of plant depends chiefly on its capacity, against their operating costs, which for a given type of plant depends chiefly on the amount of energy it generates. The amount of energy a plant generates is determined by the fraction of the time during which it operates. A plant which operates almost all the time, called a base-loaded plant, must be inexpensive to operate, but it may be worth a high capital cost because it will seldom be idle. On the other hand, a plant which operates only during peak periods of demand, called a peak-loaded plant, will often be idle and so must not be expensive to build, though it may be worth operating at high cost for short periods of time. Plants which fall between the base- and peak-loaded plants in the capital/operating cost trade-off, called cycling plants, are operated part-time, as the load cycles between base and peak levels. The distribution of a utility's generating capacity among base-loaded, cycling, and peaking plants is called its generation mix.

An important characteristic of the electric power industry is that utilities are obligated to serve whatever reasonable demands may be placed upon them by their customers. Thus utilities are very concerned with the reliability of the service they provide, and they maintain additional generating capacity beyond what is needed to meet the expected peak load to provide a reserve margin. However, because of random fluctuations in demand and random failures, or <u>outages</u>, of generating plants, utilities cannot guarantee that they will always be able to meet demand, so they must aim for a standard of service based upon some probabilistic measure of reliability. Failure to meet demand is called <u>loss of load</u>, and one of the most commonly used measures of reliability for system planning is called <u>loss-of-load probability</u>. Stochastic factors also make the cost of meeting demand uncertain, so that the utilities must use some probabilistic measure of cost in optimizing their planning decisions.

Historically, electric utility capacity planning has been based upon the twin criteria of generation mix and reserve margin. Economic considerations of minimizing cost determine the generation mix to be used, and reliability considerations determine the reserve margin to be maintained. Often the relationships between cost and generation mix and between reliability and reserve margin have been known to utility planners only through experience and heuristic rules of thumb. However, there has been considerable work done to develop more exact tools for utility capacity planning. These tools include mathematical programming planning models and probabilistic methods for measuring the reliability of electricity supply.

Mathematical programming models have been in use for capacity planning by electric utilities for about twenty-five years. An excellent survey of the models used and the extensive literature on the subject has been presented by Anderson [1]. Anderson distinguishes different kinds of models that are used at different stages of the planning process. Initial candidate plans are developed with the help of *global* models which determine the optimal capacity expansion investments. Then *simulation* models are used to determine the costs of operating each of the candidate systems. Finally, various *marginal substitutions* can be made to improve a candidate plan.

The global models used in the first stage contain within them less detailed versions of the simulation models, since in order to determine the cost of a capacity expansion plan, the operating cost must be determined. Reliability standards are handled at this stage by making *a priori* estimates of the reserve margins required.

More detailed simulation models are used in the second stage to provide more accurate estimates of the operating costs. Probabilistic methods are also used at this stage to assess reliability. There is usually considerable iteration between the global and simulation models until a candidate plan has been found which is acceptable both in cost and in reliability.

Probabilistic methods for measuring the reliability of electricity generating systems have been in use by electric utilities for more than thirty years. A brief history and bibliography can be found, for example, in Billinton [6 , 7]. Early work in this field was concerned with the methodology for calculating appropriate reliability measures and with their use for evaluating generating capacity reserve requirements. A typical use of these methods has been to determine the timing of capacity additions to a system. A reliability index is computed as a function of the system peak load, and when the load has grown to a point where the reliability is unacceptably low, the next unit on the capacity expansion schedule is added.

There are two basic approaches to the measurement of utility system reliability. The one which will be used in this thesis is the <u>loss-of-load-probability</u> method, mentioned above, which measures the expected amount of time during which the system will be unable to meet demand. The other is the <u>frequency-and-duration-of-outage</u> approach, which measures the expected time between loss of load incidents and the expected duration of these incidents.

One of the major developments in the application of probabilistic methods to capacity expansion planning was the invention of probabilistic simulation models by Baleriaux *et al.* [3] and Booth [9]. These models extend the loss of load probability method to compute not only reliability indices related to the loss of load probability but also the expected system operating costs when random plant outages can occur. Furthermore, probabilistic simulation gives an efficient recursive technique for these calculations. Using probabilistic simulation, it is possible to design mathematical programming models for planning generating capacity expansion, of the type discussed by Anderson, which explicitly take into account reliability criteria based upon probabilistic measures. The key to marrying cost minimizing capacity planning models to probabilistic reliability calculations is the use of mathematical programming decomposition techniques.

A major theme of this thesis is that efficient methods for solving the capacity expansion planning problem can be created by using decomposition techniques to exploit its natural structure. This structure arises from the inclusion of the problem of optimally operating the generating system as a subproblem within the larger problem of planning capacity expansion. In addition to their usefulness in designing efficient solution algorithms, the decomposition techniques have two other advantages. First, they lend considerable insight into the underlying economics of the capacity planning problem. Second, they encourage the design of hierarchical models with modular structure. Such a structure in a model facilitates understanding how it operates, validating it, and modifying and extending it to solve new problems.

Procedures based on decomposition methods can efficiently solve the capacity expansion planning problem because they can take advantage of a special property of the problem - when the capacities of the plants to be built are fixed at trial values, the subproblem of minimizing the operating costs can be solved very simply. As will be shown in the following chapters, the operating subproblem can often be solved in essentially closed form - no mathematical programming algorithm is required. From the solution of the operating problem, shadow prices on the trial plant capacities can be derived, representing the value of marginal changes in those capacities in changing the operating costs. These shadow prices can be used to find improved trial capacities. The trial values are generated by solving a master problem, corresponding to determination of the optimal capacity investments, in which the shadow prices are used to modify the capacity costs of the plants. An iterative scheme which alternates between solving the master problem for trial plant capacities and the subproblem for the shadow prices on these capacities can be used to converge to an optimal capacity expansion plan. These decomposition procedures are broadly applicable to the capacity expansion planning problem, and in later chapters, it will be shown how they can be used in several different formulations of the problem.

A second major theme of this thesis is the integration of mathematical programming models for capacity expansion planning with probabilistic measures of system reliability. This integration is accomplished by using probabilistic simulation as the operating subproblem of the capacity expansion model. Shadow prices associated with the trial plant capacities in the probabilistic simulation can be computed and used in a decomposition algorithm to find the optimal capacity expansion plan. The advantage of this integrated model is that a capacity expansion plan can be developed which meets explicit reliability standards, based on probabilistic measures, while minimizing expected costs.

A third major theme of this thesis is the integration of capacity expansion planning with peak-load pricing of electricity. Peak-load pricing is based on the economic concept that since the marginal cost of supplying electricity depends on the load, which varies by time of day, the price charged should also vary by time of day. There has been a great deal of research concerned with the optimal pricing of electricity, largely separate from the research on capacity planning. However, it has been widely noted that the pricing decision and the capacity planning decision are interrelated and must be made jointly. Furthermore, since peak-load pricing is rapidly moving from theoretical consideration to actual implementation, there is a need to develop pricing models which can realistically capture the complexity of the capacity planning decisions.

Decomposition techniques are natural tools for integrating the capacity planning decision with the pricing decision. As will be shown, the decomposition models easily and naturally produce the marginal cost information that is required for the pricing decision. A pricing model can be formulated in decomposition style by regarding the capacity planning model as a subproblem to compute the supply cost and formulating a master problem to compute a supply-demand equilibrium. While much work remains in formulating practical peak-load pricing models, this paper will demonstrate how the decomposition approach can be useful in their formulation.

This work is divided into three major parts. The first part introduces three mathematical programming models for utility capacity expansion planning, based on decomposition techniques. The second part discusses the technical details of the decomposition procedures for two of the models presented in the first part. Finally, the third part discusses an application of decomposition techniques to the problem of computing peak-load prices.

Part One begins with a general statement of the capacity expansion planning problem. This problem is to find a capacity expansion plan which meets given, forecast demand for electricity over a horizon of twenty to thirty years at minimum cost, which includes both the capital costs of additional generating capacity and the costs of operating that capacity. The following chapters present three specific models - a formulation based on linear programming, one based on nonlinear programming, and another using nonlinear programming which incorporates probabilistic simulation.

In Chapter 2, the capacity expansion planning problem is formulated as a linear program. It is intended to demonstrate, using a simple linear model and ordinary Benders' decomposition, techniques that will be generalized for use in the more complicated models presented in later chapters. In the linear program formulation, Benders' decomposition is used to separate out the operating subproblems from the whole capacity planning model. These subproblems, which are themselves linear programs, can be solved analytically, without requiring the use of the simplex algorithm. Given a trial capacity expansion plan, the duals of the subproblems can be solved for shadow prices, which show the marginal changes in operating cost caused by small changes in the trial plant capacities. These shadow prices are used in deriving the master problem to adjust the capacity cost coefficients. The master problem, also a linear program, is solved iteratively for the trial plant capacities. Iterations continue, alternating between the master and subproblems, until an optimal capacity plan has been found.

In Chapter 3, the capacity expansion planning problem is reformulated as a nonlinear program. The non-linear formulation is smaller than the linear formulation, because it uses a more exact representation of the problem in place of the discrete approximation used in the linear formulation. A technique called generalized Benders' decomposition is used to separate out the operating subproblems from the whole capacity expansion model. The subproblems are nonlinear, but they can still be solved analytically, and the shadow prices can be derived by applying the Kuhn-Tucker optimality conditions. The shadow prices are used in deriving the master problem, which turns out to be a linear program. Thus, the nonlinear part of the problem is confined to the subproblems, where the solutions can be obtained very simply. The explicit optimization occurs in the master problem, which is linear. As before the solution algorithm proceeds iteratively, alternately solving the master problem for a trial capacity plan and the subproblems for the shadow prices, until an optimal capacity plan has been found.

Both the linear and nonlinear programming models discussed above consider a deterministic problem in which plants are always available and there is no uncertainty about whether demand can be met. In Chapter 4, the nonlinear programming formulation is extended to consider the case in which plants can fail randomly, by including probabilistic simulation in the subproblems. In this case, it cannot be assured with certainty that demand will always be met, so it is necessary instead to specify a probable reliability with which demand will be met. Using probabilistic simulation in the subproblems allows computation of the expected costs of operating a set of trial plants and of explicit probabilistic reliability measures. The actual computation of the subproblem solution is somewhat more involved than in the previous models; however, it still does not require the use of an explicit nonlinear optimization algorithm. As before, the shadow prices on the trial plant capacities are derived from the optimality conditions for the subproblems. The operating subproblems are again separated from the capacity planning model using generalized Benders' decomposition, and the shadow prices are used to derive the master problem, which again is a linear program. Thus again, the difficult nonlinear, probabilistic part of the problem is confined to the subproblems, where it can be solved very easily. The explicit optimization is performed in the linear master program. Chapter 5 presents the results of some computational experiments with the probabilistic capacity expansion planning model.

The models discussed in Part One use a number of simplifying assumptions in order not to obscure their basic structure. Only thermal, and not hydroelectric or other nonthermal, plants have been considered. Capacity and operating costs have been represented as linear functions, and plants of any size can be built. Plant locations, transmission costs, and environmental quality standards have not been considered. Planned maintenance outages and use of spinning reserves and multiple-valve point plants have not been included. Many of these assumptions can be relaxed without disrupting the structure of the problem, and the decomposition approach can still be used. In Chapter 6 a number of extensions to the basic models are discussed which include these features.

Part Two of this thesis discusses the more technical aspects of applying generalized Benders' decomposition to the two nonlinear programming models presented in the first part. Parallel arguments are followed in Chapters 7 and 8 in developing the decomposition for the deterministic and probabilistic models, respectively. First, the generalized Benders' master problem is derived, and the solution algorithm is described. Then the solution of the subproblems and their optimal shadow prices are discussed. This discussion also includes the convexity and duality properties of the subproblem, which are required to justify the derivation of the master problem. Finally, the special situations in which the subproblems are infeasible or degenerate are discussed. Chapter 9 discusses some computational methods for the probabilistic simulation subproblem. While this problem has a fairly simple solution, the computational effort involved in calculating it could be substantial. This chapter proposes a relatively efficient technique for the computation.

Part Three of this thesis discusses the application of decomposition methods to the problem of determining peak-load prices for electricity. If the demand for electricity is varied as a parameter in the problem, the capacity expansion models presented in Part One can be regarded as generating a cost function for supplying that electricity, both the cost and the expansion plan varying as demand varies. This cost function can be embedded in a larger <u>equilibrium</u> problem in which demand is allowed to vary as a function of price. This equilibrium problem can be solved for both the prices for electricity and the capacity expansion plan. The key to solving the equilibrium problem is to use the capacity expansion model to compute the marginal costs of supplying electricity.

The approach used in this thesis differs from more traditional approaches to peak-load pricing in two respects. First, it links the determination of peak-load prices to a long-range capacity expansion planning model by using this model to generate the supply cost function. Second, it uses the probabilistic version of the capacity planning model, presented in Chapter Four of the first part. The marginal costs computed using this model differ somewhat from those used in more traditional treatments, since their time dependence is related to variations in the reliability of the supply rather than to the load level itself. Thus the model can consider the effects of random plant failures and system reliability levels on prices.

The third part begins with an introduction to the peak-load pricing problem and a brief discussion of some previous research.

Chapter 10 discusses the use of the probabilistic capacity planning model to calculate the marginal costs of supplying electricity. A method is presented for determining the contributions by components of demand at different times to the system marginal costs. This method is based on the probabilistic simulation recursive argument. It is then shown how the marginal costs are related to the dual multipliers associated with the optimal capacity plan. This relationship is discussed in more rigorous detail in Chapter 11.

Finally, Chapter 12 discusses the use of these marginal costs in an equilibrium model for computing peak-load prices. The equilibrium problem can be solved by a decomposition algorithm in which the capacity expansion model is used as a subproblem. The master problem contains the price-sensitive demand model. Trial values of the equilibrium demand are determined in the master problem and are passed to the capacity planning subproblem from which the marginal costs are calculated. The marginal costs are passed back to the master problem where they are used to compute a new trial demand estimate, and the process continues iteratively until an equilibrium has been found. This chapter concludes with a discussion of some of the practical issues involved in implementing the peak-load pricing model.

In a larger context, the models presented in this thesis demonstrate how decomposition techniques can be applied, in general, to economic planning models. Decomposition methods have several advantages which make them attractive for designing large economic models. They permit modular design of these models, in which the models are structured as essentially independent modules which communicate with one another through well-defined interfaces. Apart from these interfaces, the modules appear as "black boxes" to each other, in the sense that the internal structure of any module is of no concern to the other modules. Modularity permits hierarchical, "top-down" design of models, allowing the details of model structure to be defined by successive refinement. Furthermore, hierarchical, modular models are easier to understand and to verify, since a person need only consider the interactions of the modules at any level of the hierarchy and not their internal workings. Finally, modular models are easier to maintain and to modify, since changes can be made internally to a module without affecting the other modules. A recent trend in computer software design has been to emphasize top-down modular design using structured programming. Decomposition techniques extend these ideas to the mathematical structure of the model itself.

Decomposition models also mirror the structure of decentralized economic systems. They define interfaces between market sectors, or model modules, in terms of economic variables, prices and quantities of resources. Subject to these market indicators, each module acts to optimize its decisions. The similarity between model structure and economic structure simplifies model design and lends insight into the structure of the economic system under study.

In closing this Introduction, a word about notation. A great many indexed variables and constants appear in the following models. In order to simplify the notation, like items will often be collected together in a vector, which will be designated by the same symbol, with an underscore to indicate the vector. Thus $\underline{Y}$ is a vector consisting of the items Y^i where the index i runs from 1 to I. Indices will be represented by small letters; the upper limit of the range of an index will be represented by the same letter capitalized. Many of the models presented have a time-staged structure, for which the same variables and constraints are replicated in each time period. Time periods will generally be designated by the index t, but in showing the model for a generic period, which does not interact with the models in other periods, the index t will not be shown, but only implied, in the interest of clarity. However, when the models for different periods are brought together, the index t will be used to distinguish them. An index of notation is provided at the end of the thesis.

Finally, a distinction should be made between the terms energy and power. Power is the rate at which energy is delivered. Thus power is an instantaneous quantity. In speaking of the load on an electric power system or of the capacity of a generating plant, instantaneous power is meant. However, in speaking of operating cost, the cost of producing energy is usually intended.

Part One

Mathematical Programming Models for Electric Utility Capacity Planning

A. Organization of Part One

The purpose of Part One is to formulate three mathematical programming models for utility capacity expansion planning using decomposition techniques. The models are a linear programming formulation presented in Chapter 2, a nonlinear programming formulation presented in Chapter 3, and a nonlinear probabilistic formulation presented in Chapter 4. The underlying idea of the decomposition approach used in all of these models is to separate the capacity planning problem into two parts - a master problem, which generates trial solutions for the optimal capacity expansion plan, and subproblems, which determine the optimal operating scheme for each trial generating system. The attractive feature of the decomposition method is that when these subproblems have special structure which allows them to be solved easily, this property can be exploited in solving the larger problem in which they are embedded. The operating subproblems in each of the capacity expansion models presented here have such special structure.

The presentations of the models in the following chapters all follow the same set of steps. First, the problem of optimally operating a given set of generating plants is formulated as a mathematical program, and the special structure of its solution is discussed. It is shown how the Kuhn-Tucker optimality conditions of this problem can be used to derive shadow prices, which give the marginal changes in operating cost caused by small changes in the capacities of the plants. Second, the problem of finding an optimal generating capacity expansion plan is formulated as a mathematical program, with the operating problems embedded as subproblems. Third, it is shown how the decomposition approach can be used to develop a solution procedure for this problem. The decomposition principle is used to separate out the operating subproblems from the capacity planning model. The shadow prices are used to generate a master problem, which is another mathematical program that is solved for trial capacity expansion plans.

The algorithm developed from the decomposition principle is an iterative one. For each trial capacity expansion plan, the subproblems are solved to determine its optimal operating cost and the shadow prices on its plant capacities. These shadow prices are used to derive a new constraint in the master problem, which can then be solved again for a new trial plan. The procedure is continued, alternating between the master and subproblems, until an optimal capacity expansion plan has been found.

This part is organized so that successive models use generalizations of concepts presented in the preceding models. The linear program presented in Chapter 2 is primarily intended to demonstrate the decomposition technique using a simple linear model and ordinary Benders' decomposition. This model is reformulated as a nonlinear program in Chapter 3 and generalized Benders' decomposition is applied. In Chapter 4, the nonlinear model is extended to the probabilistic case by using probabilistic simulation in the operating subproblems. Chapter 5 presents some results of computational examples for this probabilistic model. Finally, since the models presented in the preceding chapters have used some simplifying assumptions in order not to obscure their basic structure, some extensions of the basic models are presented in Chapter 6. In many cases, the structure of the problem is not disrupted, and the decomposition procedures can still be used.

The remainder of this introduction states the capacity expansion planning problem in general form and introduces some of the notation.

B. Definitions and Problem Statement

This section presents a general statement of the problem to be discussed in the later chapters of Part One and introduces some of the concepts and definitions to be used. The problem to be considered is to find a minimum cost capacity expansion plan to meet forecast demands for electricity over a multi-year planning horizon, usually twenty to thirty years. As has been emphasized in the introductory chapter, the cost to be minimized consists of two components - the initial capital cost for the generating plants to be built and the continuing cost of operating the generating system to meet customer demand. This problem is formulated mathematically below. The formulation and notation generally follows that of Anderson [1].

Define the planning horizon as the time interval from 0 to T. Initially, any instant in this interval will be indicated by a continuous parameter, τ; however, later, it will be useful to regard the planning horizon as being made up of discrete periods (usually years or seasons) indexed by a discrete parameter t.

Let X_{jv} be the power output capacity of a plant in the system, where $j = 1,\ldots,J$ denotes the type of plant (nuclear, fossil-fueled steam, hydroelectric, gas turbine, etc.), and v indicates the vintage, or year of commissioning of the plant. (Since the planning horizon begins at time 0, use of a negative v, $-V \leq v \leq 0$, will indicate an initially available plant, so that the capacity X_{jv} is given data, while a positive v, $0 < v \leq T$, will indicate a plant yet to be built, so that X_{jv} is a decision variable.) Let C_{jv} be the present value capital cost per unit of capacity of the plant (j,v) to be built. Let $Y_{jv}(\tau)$ be the instantaneous power output of the plant at time τ, and let $F_{jv}(\tau)$ be the instantaneous cost of operation per unit of output, discounted to the present.

In order to simplify notation, it will at times be convenient to consider the capacities X_{jv} as elements of a vector $\underline{X}$. For this purpose, the indices j and v are considered combined into a single index, unique for each plant, for elements of the vector $\underline{X}$. Similarly, the $Y_{jv}(\tau)$ can be considered elements of a vector $\underline{Y}(\tau)$. Vectors and matrices of other data and variables will be introduced from time to time which are conformable with these vectors.

Let $Q(\tau)$ be the instantaneous power demand, or load, on the system at time τ. It will often be convenient to work with a discrete approximation to this time profile of demand. Period t is divided into discrete subintervals $s = 1,\ldots,S$, each of length θ_s (typically θ_s is one hour) during which the load is approximately constant at level Q_{ts}.

The discounted operating costs of plant (j,v) over the interval $\tau = 0$ to T are given by

$$\int_0^T F_{jv}(\tau)Y_{jv}(\tau)d\tau,$$

and the present value capital cost for the plant is given by

$$C_{jv}X_{jv}.$$

The objective of the planning problem is to minimize total discounted cost of building and operating the generating system:

$$\text{minimize} \sum_{j=1}^{J} \sum_{v=1}^{T} C_{jv}X_{jv} + \int_0^T \sum_{j=1}^{J} \sum_{v=-V}^{\tau} F_{jv}(\tau)Y_{jv}(\tau)d\tau.$$

Using the discrete-time approximation given above, the integral can be replaced by a sum to give

$$\text{minimize} \sum_{j=1}^{J} \sum_{v=1}^{T} C_{jv}X_{jv} + \sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{v=-V}^{t} F_{jvts}Y_{jvts}$$

where Y_{jvts} and F_{jvts} are the output and operating cost per unit output, respectively, approximated as constant during interval s of period t (assume that operating costs do not change during period t, so that $F_{jvts} = F_{jvt} \cdot \theta_s$).

There are several types of constraints on this optimization. The first is that instantaneous power output of a plant cannot exceed its capacity.

$$0 \leq Y_{jv}(\tau) \leq X_{jv} \qquad \begin{array}{l} j = 1,\ldots,J \\ v = -V,\ldots,\tau \\ \tau \in (0,T) \end{array}$$

or in discrete time

$$0 \leq Y_{jvts} \leq X_{jv} \qquad \begin{array}{ll} j = 1,\ldots,J & v = -V,\ldots,t \\ t = 1,\ldots,T & s = 1,\ldots,S \end{array}$$

The second constraint is that instantaneous power demand must be satisfied

$$\sum_{j=1}^{J} \sum_{v=-V}^{\tau} Y_{jv}(\tau) \geq Q(\tau) \qquad \tau \in (0,T)$$

or in discrete time

$$\sum_{j=1}^{J} \sum_{v=-V}^{t} Y_{jvts} \geq Q_{ts} \qquad \begin{array}{l} t = 1,\ldots,T \\ s = 1,\ldots,S \end{array}$$

where Q_{ts} is the load during interval s of period t, approximated as constant. These two, the capacity constraint and the demand constraint, are the basic constraints of the model.

Additional constraints can be included as extensions to this basic model. These extensions are discussed in more detail in Chapter 6, but a brief discussion is in order here. One extension is the inclusion of hydroelectric plants. The distinguishing characteristic of a hydroelectric plant is that the total amount of energy it can generate is limited by the amount of water stored behind the dam. Such an energy constraint has the form

$$\int_{\tau \varepsilon I} Y_{hv}(\tau) d\tau \leq H_{hv}(I)$$

where h is the plant-type index for hydroelectric plants, I is a time interval and $H_{hv}(I)$ is the amount of energy available to the plant in interval I. Usually this interval is a season or some shorter time period. In discrete form this constraint is

$$\sum_{s \varepsilon I} Y_{hvts} \leq H_{hvt}(I)$$

The amounts of hydro-energy $H_{hvt}(I)$ may be given, or they may be decision variables themselves, generating cost terms in the objective function. More complex constraints may be generated by including pumped-storage hydro facilities. Also, constraints on polluting emissions from thermal electric plants may be cast in a form similar to the energy constraints on hydroelectric plants.

Another extension of the basic model is to include additional constraints on the plant capacities, X_{jv}. It is common to require the capacity variables to be integer-valued, to represent the discrete-sized blocks in which plants can be built and the discrete site alternatives on which to build them. In addition, since electricity generating plants often exhibit economies of scale, the linear capital cost functions $C_{jv}X_{jv}$ could be replaced by concave cost functions represented piecewise-linearly, generating a mixed integer program with addition constraints.

Finally, there may be constraints called "guarantee conditions," which limit the chance that demand will not be met due to plant failures and unexpected peaks in demand. Often these conditions are approximated by multiplying the capacity of each plant by its "availability factor," to reduce its effective capacity, and requiring that the effective system capacity exceed the expected peak demand by a specified reserve margin. Such a constraint would be written

$$\sum_{j=1}^{J} \sum_{v=-V}^{t} p_{jv} X_{jv} \geq (1+m) Q_t^*$$

where p_{jv} is the availability of plant (j,v), Q_t^* is the expected peak load in period t, and m is the reserve margin. However, a more realistic representation of reliability can be formulated using probabilistic methods. These will be discussed in detail in Chapter 4.

CHAPTER 2

A LINEAR PROGRAMMING MODEL

A. Introduction

This chapter discusses formulation of the generating capacity planning problem as a linear program. The linear form allows the straightforward application of Benders' decomposition principle; decomposition of the nonlinear models of Chapters 3 and 4 is similar but somewhat more complicated. The derivation of Benders' decomposition in this chapter, using linear programming decomposition, parallels the development of the more general decomposition techniques for these nonlinear models, discussed in Part Two. Thus, this chapter provides a simple demonstration of techniques which will be used throughout this thesis.

The steps used in this chapter to develop a procedure for solving the capacity expansion planning problem as a linear program are those outlined in the introduction to Part One. In the next section, the problem of optimally operating a system of generating plants of given capacities is formulated as a linear program. This program has a simple, analytic solution (at least in the case when all plants are thermal plants with linear operating costs) called merit-order operation. Furthermore, the duality properties of this problem can be used to obtain shadow prices on the plant capacities. In the following section, the capacity expansion planning problem is formulated as a linear program, with the operating problems as subproblems. Then Benders' decomposition principle is applied to develop a solution procedure. The subproblems are separated from the main problem, and a master problem is derived using the shadow prices generated by the subproblems. The master problem is a linear program which is solved iteratively for trial capacity expansion plans. The technical details of this development are presented as a prelude to the technical development of algorithms for the nonlinear models in Part Two.

Historically, linear programming was the first mathematical optimization procedure to be applied to the utility capacity expansion problem. An excellent survey of the models used and of the extensive literature on the subject has been presented by Anderson[1]. Much of the early work on linear programming models was done at Electricité de France, presented by Massé and Gibrat[2]. Another LP model has been developed by Fernandez and Manne[3]. More recently, the LP formulation has been extended to a mixed-integer programming formulation, which more realistically represents project indivisibilities and fixed costs, by Gately[4] and by Noonan and Giglio[5]. The model of Noonan and Giglio is solved by a Benders' decomposition-based algorithm, which is a standard method for mixed-integer programs.

While Benders' decomposition has been applied before in solving these LP-based models, it apparently has not been used to exploit the special economic structure of the problem. The model presented could be extended to include integer variables to represent plants of fixed block sizes and economies of scale, as is discussed in Chapter 6; however, the primary purpose of the model presented in this chapter is to demonstrate the techniques which will be extended to the nonlinear models in later chapters.

B. Optimal Operation of the Generating System

The basic problem stated in the introduction to Part One can be solved by linear programming; however, it can have a very large number of constraints and variables. In particular, there is one variable for the operating level of each plant in each time interval, Y_{jvts}, and a corresponding capacity constraint, and there is one demand constraint for each interval s of period t in the planning horizon. Thus, for realistic problems, the linear program can be quite large, and it has been found to be expensive to solve using standard computer algorithms[6]. However, the problem has a special structure, and efficient methods for solving the problem can be developed by using decomposition techniques to exploit this structure. As has been noted above, the problem of minimizing the cost of supplying a given demand for electricity falls naturally into two components - finding a minimum-cost operating scheme for a given generating system and finding a generating system which has minimum total cost. The former problem can be cast as a subproblem which can be used in solving the latter problem.

Consider then, the subproblem of optimally operating a given set of generating plants in the subinterval s of period t. The problem is to

$$\text{minimize} \quad \sum_{j=1}^{J} \sum_{v=-V}^{t} F_{jvts} \cdot Y_{jvts} \tag{2.1}$$

$$\text{subject to} \quad \sum_{j=1}^{J} \sum_{v=-V}^{t} Y_{jvts} \geq Q_{ts} \tag{2.2}$$

$$0 \leq Y_{jvts} \leq X_{jv} \tag{2.3}$$

The solution to this problem is very simple: the plants are successively loaded up to their capacities, in order of increasing operating cost F_{jvts}, until the demand constraint (2.2) is satisfied. The last plant to be loaded, called the marginal plant, will generally not operate at full capacity.

The ordering of the plants by increasing operating costs is called the economic loading order or merit order. Thus, the optimal solution to the operating problem (2.1) - (2.3) is called merit order operation. It will be convenient to re-index the plants in merit order, and for this purpose, it is useful to define indicator constants which convert the (j,v) indices into merit order indices in period t. Note that the merit order may be different in each period, as new plants come on-line, old plants wear out, and operating costs change.

Define the indicator constant δ^{it}_{jv} equal to one if the plant whose index is (j,v) is the $i^{\underline{th}}$ plant of the merit order in period t and equal to zero otherwise. These indicators are used to pick-out the plant which holds the i^{th} position in the merit order. Define

$$X^{it} = \sum_{j=1}^{J} \sum_{v=-V}^{t} \delta^{it}_{jv} X_{jv} \qquad i = 1,\ldots,I_t$$

where I_t is the number of plants in the merit order in period t. When the plant capacities are combined into a vector $\underline{X}$, these indicators will be combined into a matrix δ_t which sorts $\underline{X}$ into merit order; thus $\delta_t \underline{X}$ gives a vector of plant capacities in merit order in period t.

It is useful, before going on, to note some properties of the indicators δ^{it}_{jv}:

i) $\sum_{j=1}^{J} \sum_{v=-V}^{t} \delta^{it}_{jv} = 1,$ only one plant can occupy the i^{th} position in the merit order;

ii) $\sum_{i=1}^{I_t} \delta_{jv}^{it} = 1,$ if $v \leq t$, a given plant (j,v) can occupy only one position in the merit order;

iii) $\delta_{jv}^{it} = 0$ for $v > t$, a plant with vintage $v > t$ has not been built yet and cannot appear in the merit order in period t;

iv) If F_{jvt} is the operating cost of plant (j,v) in period t (and operating cost doesn't depend on the time interval s), define

$$F^{it} = \sum_{j=1}^{J} \sum_{v=-V}^{t} \delta_{jv}^{it} F_{jvt}$$

Then $F^{1t} \leq F^{2t} \leq \ldots \leq F^{I_t t}$ by definition of the merit order.

v) Define

$$X^{it} = \sum_{j=1}^{J} \sum_{v=-V}^{0} \delta_{jv}^{it} X_{jv} + \sum_{j=1}^{J} \sum_{v=1}^{t} \delta_{jv}^{it} X_{jv}$$

where the first term consists of the capacities of existing plants, which are given data, and the second term consists of the capacities of plants to be built, which are decision variables.

Using the merit order notation, the operating subproblem given above, (2.1) - (2.3), can be restated as

$$\text{minimize} \quad \sum_{i=1}^{I} F^i Y^i \tag{2.4}$$

$$\text{subject to} \quad \sum_{i=1}^{I} Y^i \geq Q \tag{2.5}$$

$$0 \leq Y^i \leq X^i \qquad i = 1,\ldots,I \tag{2.6}$$

There is one such subproblem for each interval s of each period t. In order to simplify the notation, these time indices will be taken as implicit when there is no interdependence between different time intervals.

Let π be the dual multiplier, or shadow-price, associated with the demand constraint (2.5), and let λ^i be the multiplier associated with the capacity constraint in (2.6). Then the dual problem is to

$$\text{maximize} \quad Q\pi - \sum_{i=1}^{I} X^i \lambda^i \tag{2.7}$$

$$\text{subject to} \quad \pi - \lambda^i \leq F^i \qquad i = 1,\ldots,I \tag{2.8}$$

$$\pi \geq 0, \ \lambda^i \geq 0 \tag{2.9}$$

The primal problem (2.4) - (2.6) is solved by inspection, as was noted above. By the definition of the merit order, if $i < k$ then $F^i \leq F^k$. Therefore, the optimal solution is obtained by setting $Y^i = X^i$ successively in merit order until the demand constraint (2.5) is satisfied. The last plant loaded, the marginal plant, with index $i = n$, will generally not operate at full capacity. The optimal solution is

$$Y^i = \begin{cases} X^i & , \ i < n \\ Q - \sum_{i=1}^{n-1} X^i & , \ i = n \\ 0 & , \ i > n. \end{cases}$$

By complementary slackness, since $Y^i < X^i$ for $i \geq n$,

$$\lambda^i = 0, \ i \geq n.$$

Furthermore, since $Y^i > 0$ for $i \leq n$,

$$\pi - \lambda^i = F^i, \ i \leq n.$$

Hence, in particular

$$\pi = F^n$$

and therefore

$$\lambda^i = F^n - F^i, \qquad i \leq n.$$

Consider the economic meaning of the dual solution. The shadow price π on the demand constraint (2.5) is the marginal cost of increasing demand Q, which is just the cost of operating the marginal plant, the cheapest plant with slack capacity. The shadow price λ^i on the capacity constraint (2.6) is the marginal benefit of increasing the capacity of plant i. If the plant is operating at full capacity ($i < n$), then this benefit is just the difference between the cost of operating this plant and the marginal plant since adding capacity in this plant reduces the output needed from the marginal plant. If the plant is not operating at full capacity ($i \geq n$), then increasing its capacity has no value[7].

Thus, solving the primal subproblem is simply a matter of determining the marginal plant, which can be done by comparing the demand level with the capacities of the plants in merit order. Then the dual solution can be obtained by a few simple computations. The simplex algorithm is not required to solve the subproblems, and hence their solution is very efficient. This property of the subproblems can be exploited to solve the entire capacity planning problem efficiently, through the use of a decomposition procedure. The next section describes the use of Benders' decomposition principle.

C. An Optimal Generating Capacity Expansion Model and Its Solution by Benders' Decomposition

Consider the structure of the full capacity expansion planning problem in linear program form. For each interval s and period t, there is an operating subproblem (2.4) - (2.6), which can be written in the following matrix form

$$\text{minimize} \quad \underline{F}'_{ts}\underline{Y}_{ys} \tag{2.4}$$

$$\text{subject to} \quad \underline{e}'\underline{Y}_{ts} \geq Q_{ts} \tag{2.5}$$

$$0 \leq \underline{Y}_{ts} \leq \delta_t\underline{X} \tag{2.6}$$

where $\underline{Y}_{ts}$ is the vector of plant output variables in interval s of period t,

$\underline{F}_{ts}$ is the corresponding vector of operating costs;

$\underline{e}$ is a vector of ones; and

$\delta_t\underline{X}$ is the vector of plant capacities, sorted into the merit order of period t.

Then the problem of finding the minimum cost capacity expansion plan can be written

$$\text{minimize} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T}\sum_{s=1}^{S} \underline{F}'_{ts}\underline{Y}_{ts} \tag{2.10}$$

$$\text{subject to} \quad \underline{e}'\underline{Y}_{ts} \geq Q_{ts} \qquad t = 1,\ldots,T \tag{2.11}$$

$$0 \leq \underline{Y}_{ts} \leq \delta_t\underline{X} \qquad s = 1,\ldots,S \tag{2.12}$$

$$\underline{X} \geq 0 \tag{2.13}$$

where $\underline{C}$ is the vector of plant capacity costs. (For simplicity of notation, it has been assumed that all elements of the vector of plant capacities are decision variables; however, existing plants with given capacities could easily be included in the formulation.) Written out in full, this problem takes the form

$$\text{minimize} \quad \sum_{j=1}^{J}\sum_{v=1}^{T} C_{jv}X_{jv} + \sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{i=1}^{I_t} F^{its}Y^{its} \tag{2.10}$$

$$\text{subject to} \quad \sum_{i=1}^{I_t} Y^{its} \geq Q_{ts} \qquad t = 1,\ldots,T \quad s=1,\ldots,S \tag{2.11}$$

$$0 \leq Y^{its} \leq \sum_{j=1}^{J}\sum_{v=-V}^{t} \delta_{jv}X_{jv} \qquad t=1,\ldots,T \quad s=1,\ldots,S \quad i=1,\ldots,I_t \tag{2.12}$$

$$X_{jv} \geq 0 \qquad j = 1,\ldots,J \quad v = 1,\ldots,T \tag{2.13}$$

The basic idea of Benders' decomposition is to divide the problem into the two parts mentioned above, determining the optimal investments in generating capacity and determining the optimal operation of the generating system. Given a trial set of plant capacities, the operating subproblems are solved to determine the optimal operation for the system in each period, and a set of shadow prices on the plant capacities are calculated from the dual subproblems. These shadow prices are used to compute adjusted cost coefficients for the plant capacities, reflecting both their capital costs and their contribution to operating costs. The adjusted cost coefficients are used to set up constraints in a master problem, a linear program which is solved to determine a new set of trial plant capacities. These new trial capacities are inserted into the subproblems, and the solution procedure iterates in this fashion, alternating between the master and subproblems until the optimal solution is found. An important advantage of the decomposition approach is that it is often easier and more efficient to solve the master and subproblems separately than to try to solve the entire problem as a single linear program. The special properties of the subproblems allow them to be solved easily, without using the simplex algorithm, and the master problem is generally a much smaller linear program than the original problem.

In addition, at each iteration both an upper and a lower bound on the cost of the optimal solution are available, so that the algorithm can be terminated prior to optimality with known error bounds.

The following discussion demonstrates the derivation of Benders' decomposition for this problem. It is primarily intended to illustrate the principle using a relatively simple linear model, in order to motivate the discussion of the more complex decomposition methods for the nonlinear models presented in later chapters. A full discussion of Benders' decomposition principle is found in Lasdon[8].

The capacity planning problem (2.10) - (2.13) can be written in the form

$$\underset{\underline{X}\,\varepsilon\,\Omega}{\text{minimize}} \left\{\underline{C}'\underline{X} + \sum_{t=1}^{T} \sum_{s=1}^{S} \underset{\underline{Y}_{ts}\varepsilon T_{ts}}{\text{minimum}} \underline{F}_{ts}'\underline{Y}_{ts}\right\} \tag{2.14}$$

where T_{ts} is the set of all vectors $\underline{Y}_{ts}$ which satisfy constraints (2.5) and (2.6), in interval s of period t and Ω is the set of all nonnegative vectors $\underline{X}$ such that the sets T_{ts} are not empty. That is, the inner minimization is just the operating subproblem (2.4) - (2.6) discussed above. By the duality theorem of linear programming, the problem (2.14) is equivalent to

$$\underset{\underline{X}\,\varepsilon\,\Omega}{\text{minimize}} \{\underline{C}'\underline{X} + \sum_{t=1}^{T} \sum_{s=1}^{S} \underset{(\pi_{ts},\underline{\lambda}_{ts})\varepsilon\Lambda_{ts}}{\text{maximum}} \{Q_{ts}\pi_{ts} - \underline{\lambda}'_{ts}\delta_t\underline{X}\}\} \quad (2.15)$$

where $\underline{\lambda}_{ts}$ is a vector of dual multipliers on the capacity constraint in (2.6), and Λ_{ts} is the set of dual multiplier vectors $(\pi_{ts},\underline{\lambda}_{ts})$ which satisfy the constraints (2.8) and (2.9) in interval s of t. That is, the inner maximization is just the dual of the operating subproblem in interval s of t, (2.7) - (2.9).

The set of feasible solutions to the dual subproblem Λ_{ts} is a convex polyhedron which does not depend on the capacity vector $\underline{X}$, and the maximum in (2.15) is achieved at an extreme point of this set, by a well-known theorem of linear programming. Let $k = 1,\ldots,K$ index these extreme points $(\pi^k_{ts},\lambda^k_{ts})$. Then the problem (2.15) can be written

$$\underset{\underline{X}\,\varepsilon\,\Omega}{\text{minimize}} \{\underline{C}'\underline{X} + \sum_{t=1}^{T} \sum_{s=1}^{S} \underset{k=1,\ldots,K}{\text{maximum}} \{Q_{ts}\pi^k_{ts} - \underline{\lambda}^k_{ts}\delta_t\underline{X}\}\} \quad (2.16)$$

In order to insure that the primal subproblems have feasible solutions (that is, the sets T_{ts} are not empty), it is sufficient to require that enough capacity be built to meet the peak demand Q^*_t in each period t. This requirement takes the form of the constraints

$$\sum_{i=1}^{I_t} X^{it} \geq Q_t^* \qquad t = 1,\ldots,T$$

or in vector form

$$\underline{e}'\delta_t \underline{X} \geq Q_t^* \qquad t = 1,\ldots,T$$

where $Q_t^* = \underset{s=1,\ldots,S}{\text{maximum}} \; Q_{ts}$.

The set Ω consists, at most, of capacity vectors satisfying these constraints. Additional constraints on the capacities imposed in the original problem may also be included in defining Ω.

The capacity expansion planning problem can now be written from (2.16) in the following form

$$\underset{Z,\underline{X}}{\text{minimize}} \quad Z \tag{2.17}$$

$$\text{subject to } Z \geq (\underline{C}' - \sum_{t=1}^{T} \sum_{s=1}^{S} \underline{\lambda}_{ts}^{k} \delta_t)\underline{X} + \sum_{t=1}^{T} \sum_{s=1}^{S} Q_{ts}\pi_{ts}^{k} \tag{2.18}$$

$$k = 1,\ldots,K$$

$$\underline{e}'\delta_t \underline{X} \geq Q_t^* \qquad t = 1,\ldots,T \tag{2.19}$$

$$\underline{X} \geq 0$$

and any other constraints on $\underline{X}$ imposed in the original problem.

This program is the master problem.

The master problem is actually solved by successively generating the constraints (2.18). Starting with an initial trial set of capacities $\underline{X}^1$, the subproblems (2.4) - (2.6) are solved for the dual multipliers to generate the first (k=1) Benders' cut, as the constraints (2.18) are called. In general, a relaxed master problem consisting of constraints (2.18) with $k = 1,\ldots,\ell-1$ is solved for a new trial capacity plan $\underline{X}^\ell$. The subproblems are then solved with these capacities, and the associated shadow prices $(\pi^\ell_{ts},\underline{\lambda}^\ell_{ts})$ are used to generate the next Benders' cut, with $k = \ell$. Notice that these shadow prices are used in (2.18) to adjust the cost coefficients for the plant capacities. The new Benders' cut is, in a sense, the "most violated" of the constraints (2.18) not yet included in the master problem. Since the current set of Benders' cuts in the relaxed master problem is a subset of the entire set of such constraints, the value Z^ℓ generated by solving the master problem at each iteratiion ℓ is a lower bound on the cost of the optimal solution. However, if the current trial values Z^ℓ and $\underline{X}^\ell$ satisfy the newly generated Benders' cut (k=ℓ), the current solution is, in fact, optimal. If not, then the value of the new constraint with the current value $\underline{X}^{\ell}$

$$(C' - \sum_{t=1}^{T} \sum_{s=1}^{S} \underline{\lambda}_{ts}^{\ell} \delta_t) \underline{X}^{\ell} + \sum_{t=1}^{T} \sum_{s=1}^{S} Q_{ts} \pi_{ts}^{\ell}$$

is an upper bound on the cost of the optimal solution, since the constraints (2.19) guarantee that the trial solution $\underline{X}^{\ell}$ is feasible.

Though the model presented in this chapter is concerned primarily with satisfying the demand and capacity constraints, (2.11) and (2.12), the Benders' algorithm is well-suited to treat extensions to this basic model which include additional constraints. Additional constraints on the capacity variables $\underline{X}$ can be incorporated into the master problem. Even integer constraints on the capacities can be included, since Benders' decomposition is a standard method for solving mixed-integer programs. In this case, however, the master problem is no longer a linear program. Additional constraints on the generator output variables $\underline{Y}_{ts}$ can be incorporated into the subproblems. Some types of these constraints are compatible with the special structure, so that the subproblems could still be solved without explicit use of a mathematical programming algorithm. However, even if this is not the case, the decomposition approach simplifies the solution of the entire problem by breaking it up into smaller pieces. Further discussion of extensions of the basic model are found in Chapter 6.

CHAPTER 3

A NONLINEAR PROGRAMMING MODEL FOR THE DETERMINISTIC CASE

A. Introduction

In this chapter, the generating capacity planning problem is formulated as a nonlinear program. The nonlinear formulation offers two advantages over the linear formulation discussed in the previous chapter. First, the nonlinear version is more compact than the linear formulation. The large size of the linear program formulation is chiefly a result of using a discrete-time approximation to the load profile, Q_{ts}. Thus, there must be one operating subproblem for each subinterval s in period t. The nonlinear formulation uses a continuous representation of the load in each period t, called a load duration curve, and uses explicitly the optimality of merit order operation in formulating a single subproblem for each period t. The subintervals s used in the discrete-time approximation are not needed in the nonlinear formulation. The second advantage is that the nonlinear formulation can be directly extended to the probabilistic case which will be discussed in Chapter 4. This extension is possible because there is a direct analogy between the nonlinear subproblems which will be formulated in this chapter and probabilistic simulation which will be used in the next chapter.

The development of a solution procedure for the nonlinear formulation in this chapter follows the general steps described in the introduction to Part One. The next section discusses the formulation and solution of the operating subproblems. The formulation uses explicitly the fact that merit order operation is optimal, as shown in the preceding chapter, so that the solution is trivial. However, the Kuhn-Tucker conditions of the problem can be used to obtain the shadow prices on the plant capacities, which is extremely useful information. In the following section, the capacity expansion problem is formulated as a mathematical program using the operating problems as subproblems. The use of decomposition to solve this problem is described. Since the problem is nonlinear, the generalized Benders' decomposition of Geoffrion[1] must be used instead of the ordinary Benders' decomposition that was used for the linear model. The technical details of applying generalized Benders' decomposition to this problem are discussed in Chapter 7.

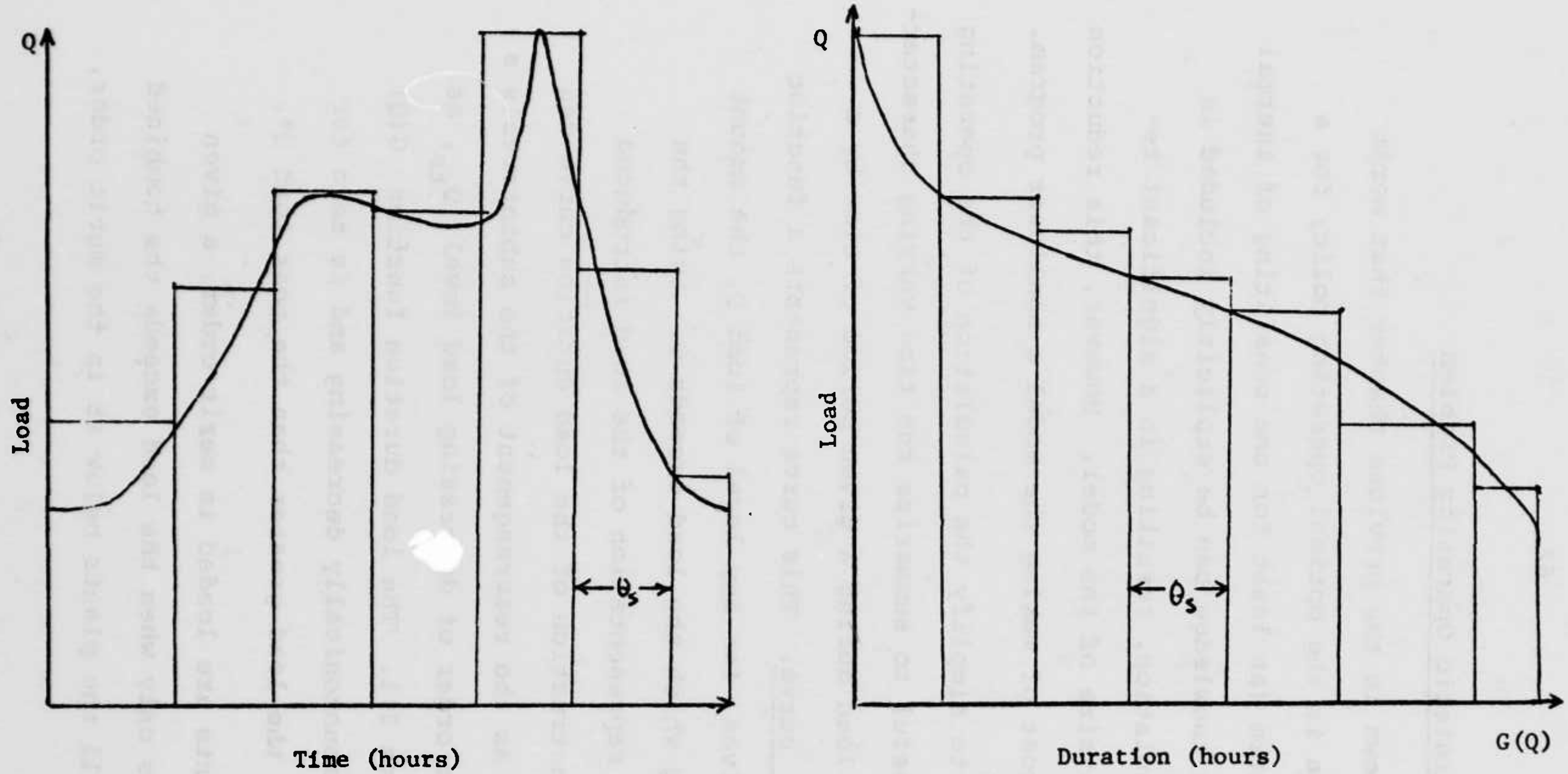
Historically, the nonlinear programming models for utility capacity expansion planning grew out of the linear models in an attempt to reduce the problem to computationally manageable size. Work on such models has been presented by a number of authors. Phillips *et al*[2] have presented a nonlinear model which is similar in many respects to the model presented in this chapter. They have developed a solution procedure based on nonlinear programming dual multipliers. Another such model is described by Bessière[3]. A third model, presented by Beglari and Laughton[4], separates the problem into a capacity expansion planning linear program and operating subproblems, as is proposed in this chapter, but uses plant capacity factors, rather than shadow prices, to link the models together. Apparently, the decomposition approach presented here has not been considered before.

B. The Deterministic Operating Problem

It was shown in the previous chapter that merit order operation is the optimal operating policy for a generating system (at least for one consisting of thermal plants). This knowledge can be explicitly included in the model formulation, resulting in a significant reduction in the size of the model. However, this reduction comes at the cost of making the model a nonlinear program.

In order to simplify the calculation of the operating cost, it is useful to summarize the time varying characteristics of the load during a given period of time by a load duration curve. This curve represents a function $G(Q)$ which gives, for any level of load Q, the amount of time during which the load exceeds Q. Using the discrete-time representation of the load introduced above, the construction of the load duration curve can be visualized as the rearrangement of the subintervals s of period t in order of decreasing load level Q_{ts}, as shown in Figure 3.1. The load duration function $G(Q)$ is therefore monotonically decreasing and is zero for all levels of the load greater than the peak load Q^*.

When plants are loaded in merit order, a given plant operates only when the load exceeds the combined capacity of all the plants below it in the merit order,



Daily Load Profile

Daily Load Duration Courve

Figure 3.1
Derivation of the Load Duration Curve

and then it operates either at capacity or at the level of excess, whichever is smaller. Thus if plants are "stacked" under the load duration curve in merit order, the energy generated by each plant is given by the area it "cuts out" under the load duration curve (see Figure 3.2).

More formally, define the cumulative capacity of all plants up to and including plant i as

$$U^i = \sum_{n=1}^{i} X^n \qquad i = 1,\ldots,I$$

or recursively,

$$U^i - U^{i-1} = X^i \qquad i = 1,\ldots,I \tag{3.1}$$

where $U^0 = 0$.

The load level U^{i-1} is called the loading point of plant i, since plant i begins generating when the load reaches U^{i-1}.

The amount of energy generated by the $i^{\underline{th}}$ plant in the merit order in a given time period is

$$\int_{U^{i-1}}^{U^i} G(Q)dQ,$$

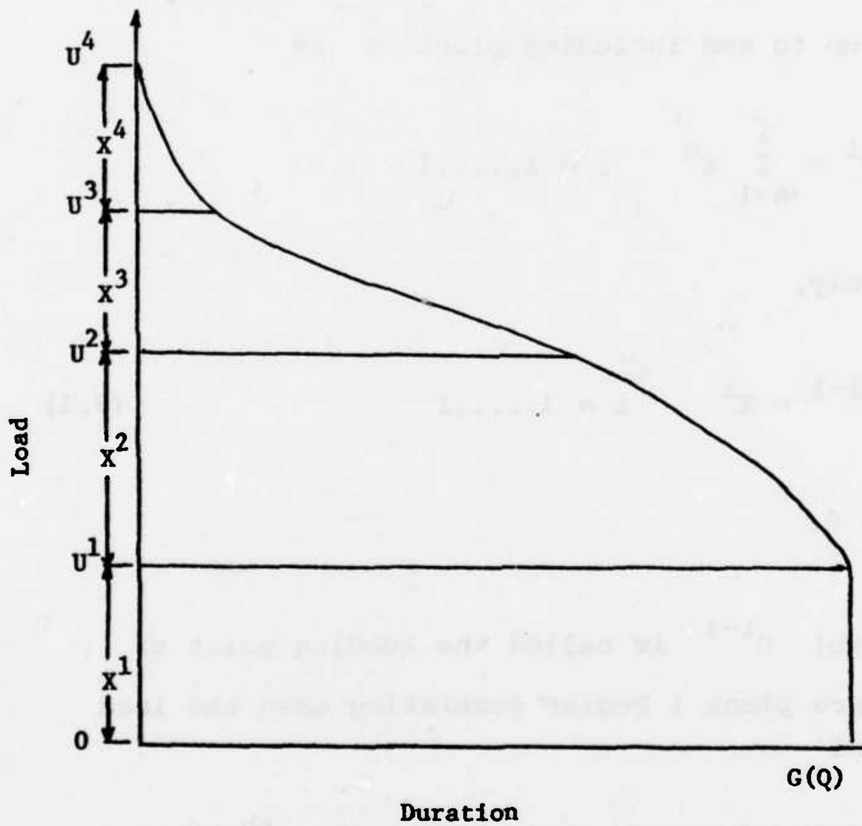


Figure 3.2
Merit Order Operation

and the cost of operating this plant is thus

$$F^i \int_{U^{i-1}}^{U^i} G(Q)\,dQ,$$

where F^i is the operating cost per unit of energy produced by plant i. Therefore, the optimal cost of operating the generating system in a given period is

$$\sum_{i=1}^{I} F^i \int_{U^{i-1}}^{U^i} G(Q)\,dQ. \qquad (3.2)$$

In order for this operating scheme to be feasible, the system must have sufficient capacity to meet the peak load; hence

$$U^I \geq Q^* \qquad (3.3)$$

The problem of minimizing the operating cost (3.2) subject to the loading order constraints (3.1) and the peak load constraint (3.3) can be regarded as the operating subproblem in period t, analogously to the subproblems defined in the previous chapter:

$$\text{minimize} \quad \sum_{i=1}^{I} F^i \int_{U^{i-1}}^{U^i} G(Q)\,dQ \qquad (3.2)$$

$$\text{subject to} \quad U^i - U^{i-1} = X^i \qquad i = 1,\ldots,I \tag{3.1}$$

$$U^I \geq Q^* \tag{3.3}$$

$$U^i \geq 0 \qquad i = 1,\ldots,I$$

(Clearly, the load duration curve is defined for a given period t, and the merit order, peak load, and cost coefficients all depend on t as well. However, the index t has been suppressed here for clarity of notation. It will be used below where needed.) As before, the plant capacities X^i are considered constants in the subproblem.

Of course, optimization of these subproblems is trivial, because only one of two situations can occur. If the peak-load constraint (3.3) is satisfied, the loading order constraints (3.1) yield a single solution, which is optimal. If the peak-load constraint is not satisfied, the subproblem is infeasible. However, it is useful to regard this subproblem as an optimization in order to compute shadow prices on the plant capacities.

The objective function (3.2) is separable, and it is convex on the feasible region defined by constraints (3.1) and (3.3), as will be shown in Chapter 7. Thus a necessary and sufficient condition for optimality in the subproblem is the existence of a set of dual multipliers satisfying the Kuhn-Tucker conditions. Define λ^i and π as the dual multipliers, or shadow prices, on the $i^{\underline{th}}$ loading order constraint (3.1) and on the peak-load constraint (3.3) respectively. Because (3.1) are equality constraints, λ^i is unrestricted in sign while π must be non-negative. Assuming that the subproblem is feasible and that $U^i > 0$ for all i, the Kuhn-Tucker conditions give the set of equations

$$\begin{aligned} \lambda^i - \lambda^{i+1} &= (F^{i+1} - F^i)\, G(U^i) \qquad i = 1,\ldots,I-1 \\ \lambda^I - \pi &= 0 \end{aligned} \tag{3.4}$$

If, as will often be the case, $U^I > Q^*$, then $\pi = 0$ and the shadow prices λ^i can be determined by solving the set (3.4) by backward recursion. (The cases when $U^I = Q^*$ or $U^i = 0$ for some i are degenerate cases which are discussed in Chapter 7.) The shadow price λ^i represents the marginal operating cost reduction resulting from an increase in the plant capacity X^i. The shadow price π represents the marginal cost of meeting additional peak demand Q^*.

Thus, as was true for the linear programming subproblem discussed in Chapter Two, this nonlinear operating subproblem can be solved very easily by determining the plant loading points U^i. Then the dual solution can be obtained by a simple computation. It is not necessary to use a mathematical programming algorithm, and hence, the solution of the subproblems can be very efficient. This special structure of the subproblems can be exploited in designing an efficient procedure to solve the entire capacity planning problem through the use of a decomposition technique. The next section discusses the application of generalized Benders' decomposition.

C. The Capacity Expansion Planning Problem and Its Solution by Generalized Benders' Decomposition

Consider the structure of the capacity expansion planning model using the nonlinear operating cost model. There is an operating subproblem, described in the previous section, for each period t, which, in order to simplify notation, can be written in the following matrix form

$$\text{minimize} \quad F_t(\underline{U}_t) \tag{3.5}$$

$$\text{subject to} \quad M_t\underline{U}_t = \delta_t\underline{X} \tag{3.6}$$

$$N_t\underline{U}_t \geq Q_t^* \tag{3.7}$$

$$\underline{U}_t \geq 0$$

where $\underline{X}$ is the vector of plant capacities X_{jv},

$\underline{U}_t$ is the vector of plant loading points, U^{it}, in period t, and

Q_t^* is the peak load in period t.

Then the objective function for the subproblem (3.2) in period t is given by the function $F_t(\underline{U}_t)$, the loading order constraints (3.1) are represented by the matrix M_t, the peak load constraint (3.3) is represented by the vector N_t, and the matrix δ_t sorts the vector of plant capacities into merit order, as described previously.

The problem of finding a minimum cost capacity expansion plan can be written as

$$\text{minimize} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} F_t(\underline{U}_t) \tag{3.8}$$

$$\text{subject to} \quad M_t\underline{U}_t = \delta_t\underline{X} \qquad t = 1,\ldots,T \tag{3.9}$$

$$N_t\underline{U}_t \geq Q_t^* \tag{3.10}$$

$$\underline{X} \geq 0, \; \underline{U}_t \geq 0$$

where $\underline{C}$ is the vector of plant capacity costs. (Again, for simplicity of notation, it has been assumed that all elements of the capacity vector $\underline{X}$ are decision variables; however, existing plants with given capacities could easily be included in the formulation.) Written out in full, this problem takes the form

$$\text{minimize} \quad \sum_{j=1}^{J} \sum_{v=1}^{T} C_{jv}X_{jv} + \sum_{t=1}^{T} \sum_{i=1}^{I_t} F^{it} \int_{U^{i-1,t}}^{U^{it}} G_t(Q)\,dQ \tag{3.8}$$

$$\text{subject to} \quad U^{it} - U^{i-1,t} = \sum_{j=1}^{J} \sum_{v=1}^{T} \delta_{jv}^{it} X_{jv} \quad i=1,\ldots,I_t \tag{3.9}$$

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

$$U^{I_t t} \geq Q_t^* \qquad t = 1,\ldots,T \tag{3.10}$$

$$X_{jv} \geq 0 \quad U^{it} \geq 0$$

Define $\underline{\lambda}_t$ as a vector of dual multipliers associated with each set of loading order constraints (3.9) and π_t as a dual multiplier associated with each peak-load constraint (3.10). Then the Kuhn-Tucker optimality conditions for the capacity planning problem (3.8) - (3.10) give the following conditions

$$\begin{aligned} \lambda^{it} - \lambda^{i+1,t} &= (F^{i+1,t} - F^{it})G_t(U^{it}) \quad i = 1,\ldots,I_t - 1 \\ & \qquad\qquad t = 1,\ldots,T \\ \lambda^{I_t t} - \pi_t &= 0 \end{aligned} \tag{3.11}$$

(assuming $U^{it} > 0$ and $U^{I_t t} > Q_t^*$)

and

$$C_{jv} - \sum_{t=1}^{T} \sum_{i=1}^{I_t} \lambda^{it} \delta_{jv}^{it} \geq 0 \quad \text{for all } j,v \tag{3.12}$$

with equality if $X_{jv} > 0$.

(The degenerate cases when $U^{it} = 0$ or $U^{I_t t} = Q_t^*$ are discussed in Chapter 7.) The equations (3.11) are just the equations (3.4) derived in the previous section. As noted before, they can be easily solved by backward recursion for the shadow prices λ^{it}. These shadow prices are used in (3.12) to "price out" the capacity varaibles X_{jv}. If the cost of building a plant exceeds the benefits derived from operating it ((3.12) holds with strict inequality) then the plant will not be built ($X_{jv} = 0$).

A solution procedure can be suggested along the lines of the Benders' decomposition of the previous Chapter 2. Because of the nonlinear structure problem, the generalized Benders' decomposition of Geoffrion [17] must be used. However, because the cost function and the constraints are separable in X_{jv} and U^{it}, it turns out that the master problem is a linear program. Though the subproblems are nonlinear, they are solved by inspection as discussed above. The basic idea in deriving generalized Benders' decomposition is the same as that used in the derivation of ordinary Benders' decomposition in the previous chapter, and the technical details are discussed in Chapter 7.

The master problem, derived according to the discussion of Chapter 7, is

$$\underset{Z,\ \underline{X}}{\text{minimize}} \quad Z \tag{3.13}$$

$$\text{subject to } Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [F_t(\underline{U}_t^k) + \underline{\lambda}_t^k \delta_t (\underline{X}^k - \underline{X})] \quad k=1,\ldots,K \tag{3.14}$$

$$\underline{e}'\delta_t \underline{X} \geq Q_t^* \qquad t = 1,\ldots,T \tag{3.15}$$

$$\underline{X} \geq 0$$

and any additional constraints imposed on the $\underline{X}$ in the original problem.

The index k refers to trial solutions of the master and subproblems generated at iteration k. The constraints (3.15) insure feasibility in the subproblems. They can be written in component form as

$$\sum_{i=1}^{I_t} X^{it} \geq Q_t^* \qquad t = 1,\ldots,T \tag{3.15}$$

in which it can be seen that they require that sufficient capacity be built to meet the peak demand in each period t.

As before, the master problem is solved by successively generating the constraints (3.14). Starting with an initial trial set of capacities $\underline{X}^1$, the subproblems (3.5) - (3.7) are solved for each period, and the shadow prices are calculated using (3.11) to generate the first (k=1) Benders' cut (3.15). In general, a relaxed master problem consisting of constraints (3.15) with $k = 1,\ldots,\ell-1$ is solved for a new trial capacity plan $\underline{X}^{\ell}$. The subproblems are then solved with these capacities, and the associated shadow prices $(\pi_t, \underline{\lambda}_t)$ are used to generate the next Benders' cut, with $k = \ell$. The algorithm proceeds iteratively, alternating between the master problem and the subproblems until optimality is achieved. Note that the use of the nonlinear functions $G_t(Q)$ is confined to the subproblems where no explicit optimization is performed. The subproblems can be regarded as "black boxes" which take the trial capacities X_{jv} as inputs and produce the shadow prices λ^{it} as outputs and which could be called as subroutines in the optimization algorithm. The explicit optimization occurs in the master problem, which is a linear program. A detailed discussion of the technical details of the solution algorithm using generalized Benders' decomposition is found in Chapter 7.

CHAPTER 4

A MODEL FOR THE PROBABILISTIC CASE

A. Introduction

This chapter discusses a formulation of the generating capacity expansion problem which explicitly considers reliability standards defined by probabilistic measures. The model is an extension of the nonlinear program developed for the deterministic case in the previous chapter. This extension is accomplished by using the technique of probabilistic simulation[1], which calculates the impact of random plant failures on operating costs and on ability to serve demand. As will be shown, probabilistic simulation generates operating subproblems which are directly analogous to the subproblems of the deterministic model in the previous chapter.

The steps used to develop the probabilistic model in this chapter parallel those used to develop the models presented in the preceding two chapters. First, probabilistic simulation is explained and used to set up the operating problem as a nonlinear program. Expressions are derived for the expected system operating cost and for two probabilistic reliability measures, loss-of-load probability and expected unserved energy. The suitability of each for defining reliability standards is compared. Shadow prices on the plant capacities are derived from the Kuhn-Tucker conditions for the subproblem. Next, the capacity expansion planning problem is formulated as a mathematical program, with these operating problems incorporated as subproblems. Then generalized Benders' decomposition is applied to develop a solution procedure for this problem. As before, the subproblem shadow prices are used to derive the master problem, which, in this case as before, is a linear program. The technical details of developing the solution procedure are found in Chapter 8; those of computing the shadow prices in Chapter 9. The chapter closes with a discussion of an alternative treatment of reliability in the planning model using costs rather than constraints.

Probabilistic methods have been in use for a long time in evaluating power system reliability[2]. The development of probabilistic simulation by Baleriaux et al. and Booth[3] marked a major advance by providing a relatively efficient method for computing widely used reliability measures and operating costs under probabilistic conditions. An attempt to integrate probabilistic simulation into mathematical programming models for capacity expansion planning has been made in the Generation Expansion Model (GEM) of Schweppe et al.[4] In this model, probabilistic simulation is used as an operating subproblem, and the capacity expansion model is a linear program. Communication between the subproblems and the linear program is achieved using plant capacity factors, rather than shadow prices, and convergence difficulties have been encountered.[5] A similar approach has been taken by Beglari and Laughton[6]. Telson[8] used the GEM model in his study of the costs and benefits of changing electricity-supply reliability levels.

The approach used in this chapter, using shadow prices to interface the probabilistic simulation subproblems with the mathematical program for capacity planning, has apparently not been used before. This approach allows the rigorous development of solution algorithms based on decomposition theory for mathematical programs.

B. Probabilistic Simulation and the Operating Problem

Consider the effects of random plant failures, or outages, on the operation of the generating system. A plant outage has two effects - first, it reduces the total amount of energy the plant produces over a given time period and second, it causes the plants above it in the merit order to produce more energy, at higher cost, in order to compensate. In addition, it is possible that enough capacity will be down at some time that demand cannot be satisfied, a condition known as loss of load.

The operation of a plant subject to random failures is often regarded as an alternating renewal process[7], which is a stochastic process consisting of alternating periods of operation and outage (see Figure 4.1). The time spent in operation before a failure, m, and the time spent in repair before a return to service, r, are randomly drawn, independently from two different distributions, and each failure-repair cycle is independent of, but probabilistically identical to, the others. In steady state, the probability of finding the plant in operation at an arbitrary moment is

$$p = \frac{\bar{m}}{\bar{m} + \bar{r}}$$

where the overbars indicate the means of the respective

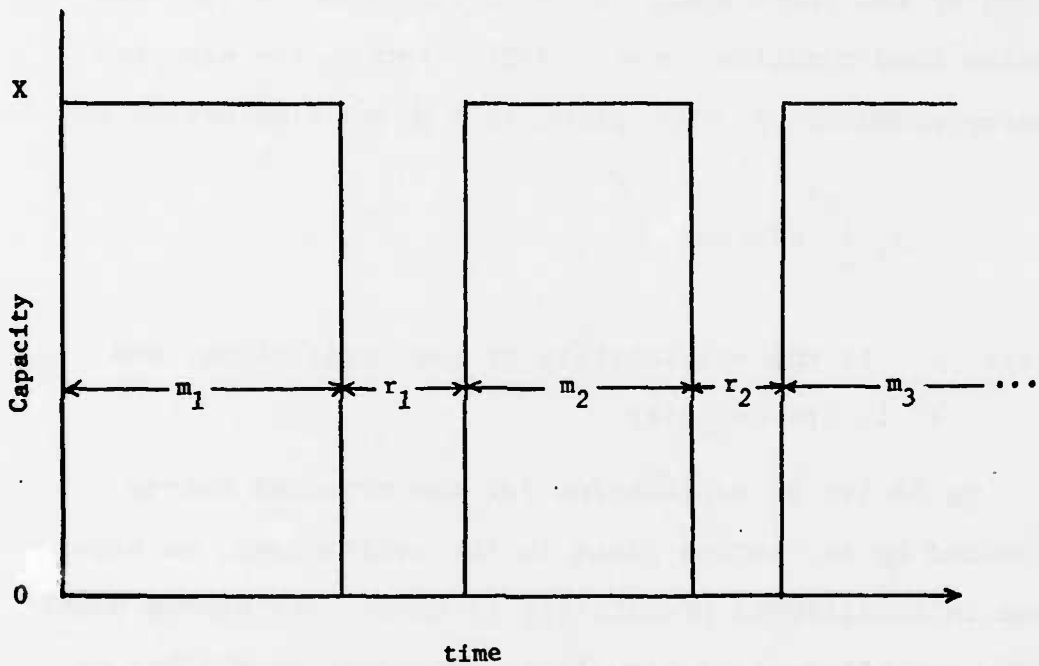


Figure 4.1
A Typical History of Plant Operation

random variables. This probability p is called the availability of the plant. In the following discussion, its complement $q = 1-p$ will also be used.

Given that it operates, the load duration curve faced by the first plant in the merit order is just the system load duration curve, $G(Q)$. Hence, the expected energy produced by this plant in a given time period is

$$p_1 \int_0^{X^1} G(Q)dQ$$

where p_1 is the availability of the first plant, and X^1 is its capacity.

To derive an expresssion for the expected energy produced by the second plant in the merit order, an argument in conditional probability is used. The second plant faces two alternative situations depending on whether or not the first plant is operating. Given that the first plant is operating, the second plant is loaded after the first and produces expected energy

$$p_2 \int_{X^1}^{X^1+X^2} G(Q)dQ.$$

This situation occurs with probability p_1. Given that the first plant is not operating, the second "drops down" in the merit order and produces expected energy

$$p_2 \int_0^{X^2} G(Q)dQ.$$

This situation occurs with probability $q_1 = 1-p_1$. Thus, the expected energy produced by the second plant is

$$p_2(p_1 \int_{X^1}^{X^1+X^2} G(Q)dQ + q_1 \int_0^{X^2} G(Q)dQ)$$

or

$$p_2 \int_{X^1}^{X^1+X^2} \{p_1G(Q)dQ + q_1G(Q-X^1)\}dQ,$$

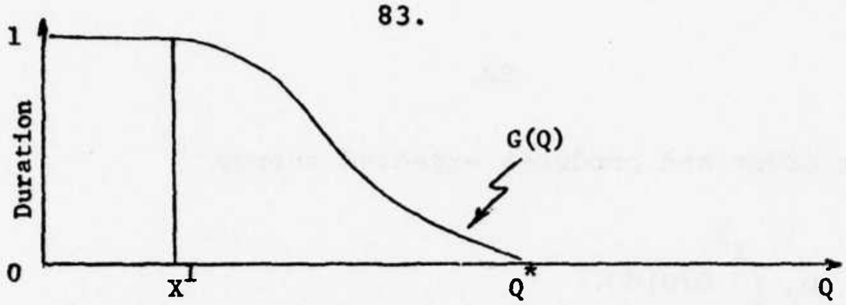
where the term in brackets is the equivalent load duration curve faced by the second plant (see Figure 4.2).

Using this same conditional probability argument for each successive plant, the equivalent load duration curve for the $i + 1^{\underline{st}}$ plant in the merit order is defined by

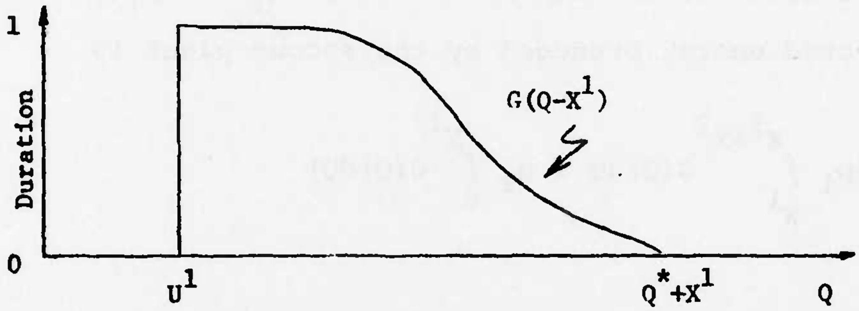
$$G_{i+1}(Q) = p_iG_i(Q) + q_iG_i(Q-X^i) \quad i = 1,...,I \tag{4.1}$$

where $G_1(Q) \equiv G(Q)$.

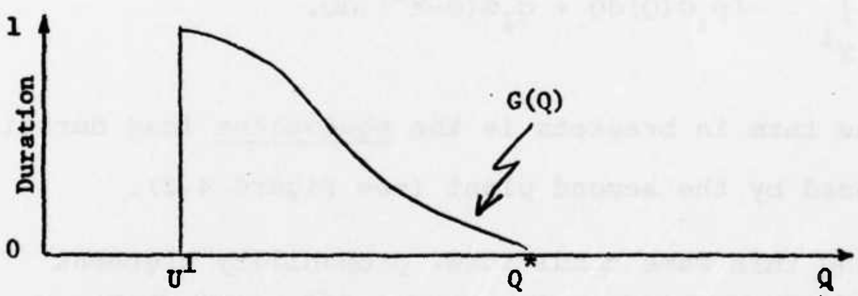
This recursive relationship is known as probabilistic simulation. The expected energy delivered by the $i^{\underline{th}}$ plant



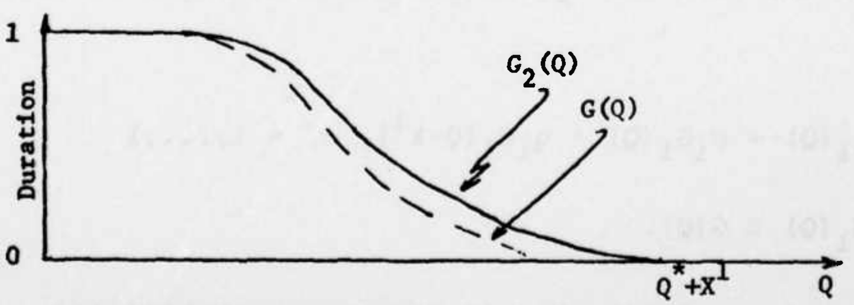
1. The first plant faces the system load duration curve.



2. If the first plant fails, the second plant faces the entire system load duration curve. This event has probability q_1.



3. If the first plant operates, the second plant faces only the load not served by the first. This event has probability p_1.



4. The equivalent load duration curve faced by the second plant is the sum of these two curves weighted by their probabilities.

Figure 4.2
Derivation of the Equivalent Load Duration Curve

(Adapted from Finger [13].)

in the merit order during the given time period is then

$$p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ$$

where, as before, U^i is the cumulative capacity defined by

$$U^i - U^{i-1} = X^i \qquad i = 1,\ldots,I \tag{4.2}$$

with $U^0 = 0$.

The expected cost of operating the system during the given period is

$$\sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ. \tag{4.3}$$

Note the similarity of this expression to the corresponding cost expression in the deterministic case (3.2).

Since each plant has a non-zero probability of failure, it is not possible to guarantee with certainty that demand will always be satisfied. It is possible that enough plants will have failed at one time that the load will exceed the available capacity. The likelihood of this occurring is often measured by the <u>loss-of load probability</u> (LOLP), which is defined as the expected amount of time during which the load will exceed the available capacity during a given time period. The LOLP is given by

$$\text{LOLP} = G_{I+1}(U^I)$$

since the equivalent load duration curve $G_{I+1}(Q)$ represents the load remaining to be served after all the plants have been loaded (see Figure 4.3). This measure is called a "probability" because the load duration curve is often regarded as being analogous to a probability distribution for demand.

The loss-of-load probability can be used as a reliability criterion for design. Instead of the peak-load constraint (3.3), which is no longer a useful standard, a reliability constraint of the following form is used:

$$G_{I+1}(U^I) \leq \varepsilon \tag{4.4}$$

where ε represents the desired reliability level (a typical design target is a LOLP of one day in ten years).

It has been argued (see, for example, Telson[8]) that the loss-of-load probability is not an entirely satisfactory measure of reliability, since it takes account only of the likelihood that some load will not be met and not of the

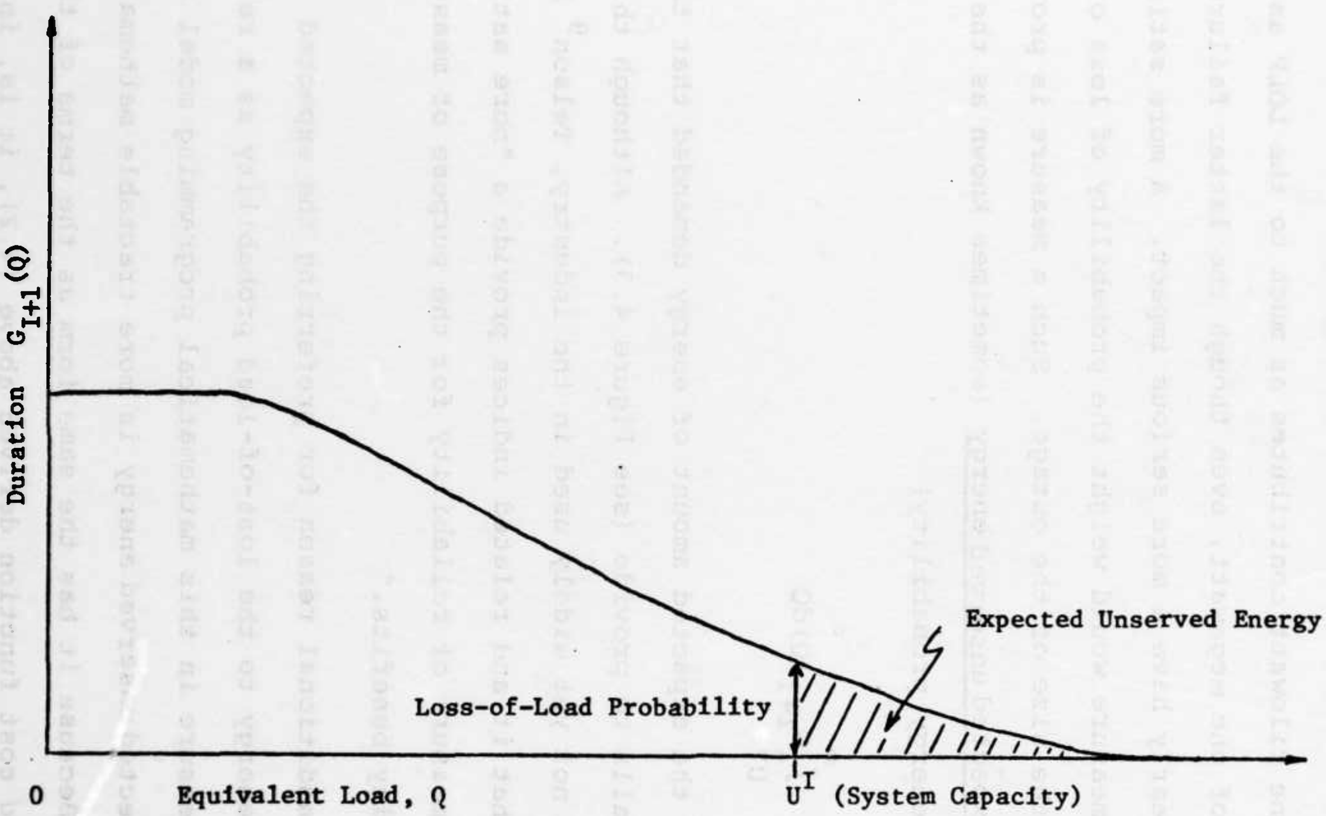


Figure 4.3
Loss-of-Load Probability and
Expected Unserved Energy

size of the deficit. For example, failure to meet the load by one kilowatt contributes as much to the LOLP as a failure of one megawatt, even though the latter failure would clearly have a more serious impact. A more satisfactory measure would weight the probability of loss of load by the size of the outage. Such a measure is provided by the expected unserved energy (sometimes known as the loss-of-energy probability)

$$\int_{U^I}^{\infty} G_{I+1}(Q)\,dQ$$

which is the expected amount of energy demanded that the system fails to provide (see Figure 4.3). Although this index is not yet widely used in the industry, Telson[9] has argued that it and related indices provide a "more satisfactory measure of reliability for the purpose of measuring reliability benefits."

An additional reason for preferring the expected unserved energy to the loss-of-load probability as a reliability measure in this mathematical programming model is that expected unserved energy is more tractable mathematically. Because it has the same form as the terms of the operating cost function derived above (4.2), it is, in a certain sense, "compatible" with them. This compatibility will be used in Chapters 8 and 9 in proving convexity properties for the problem and in computing shadow prices. While LOLP (or other probabilistic reliability measures) could be used in this planning model, the expected unserved energy will be used here because of these two advantages it has over the LOLP.

Thus a reliability constraint on expected unserved energy will be imposed

$$\int_{U^I}^{\infty} G_{I+1}(Q)\,dQ \leq \varepsilon \qquad (4.5)$$

where ε represents the desired reliability standard.

The problem of minimizing the expected operating cost subject to the reliability constraint is the operating subproblem in this probabilistic case, analogously to the deterministic operating subproblem discussed in the previous chapter. In this probabilistic case, however, the structure of the problem is somewhat more complicated. Since minimum cost is achieved when the reliability constraint is exactly satisfied, some of the plants may not be used at full capacity in the optimal solution. Let Y^i be the utilization level of the i^{th} plant in the merit order, where

$$0 \leq Y^i \leq X^i.$$

Usually this utilization level will be set equal to capacity. However, if the system has excess capacity, more than is required to satisfy the reliability constraint, some of the more expensive plants high in the merit order may be shut down. Their utilization levels would be set to zero.

The operating subproblem for the probabilistic model can then be stated as

$$\text{minimize} \quad \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)\,dQ \tag{4.6}$$

$$\text{subject to} \quad \int_{U^I}^{\infty} G_{I+1}(Q)\,dQ \leq \varepsilon \tag{4.7}$$

$$0 \leq Y^i \leq X^i \tag{4.8}$$

where the plant loading points are now defined by

$$U^i - U^{i-1} = Y^i \qquad i = 1,\ldots,I \tag{4.9}$$

with $\qquad U^0 = 0$

and the equivalent load duration functions, written to show their dependence on Y^i explicitly, are

$$G_{i+1}(Q;Y^1,\ldots,Y^i) = p_i G_i(Q;Y^1,\ldots,Y^{i-1}) + q_i G_i(Q-Y^i;Y^1,\ldots,Y^{i-1}) \quad i=1,\ldots,I$$

with $G_1(Q) = G(Q)$. (As before, though the system load duration curve, operating cost coefficients, and merit order all depend on the period t, this index has been suppressed for clarity of notation. It will be introduced below, where needed.) Again the plant capacities X^i are considered constant data in the subproblems.

The optimal solution to this subproblem is intuitively simple: Set $Y^i = X^i$ successively in merit order until the unserved energy constraint (4.7) is exactly satisfied. The last plant so loaded, the marginal plant, will generally not have to be used to capacity. The plants above the marginal plant will not be used. Let n be the merit order index of the marginal plant; then this solution can be written

$$Y^i = \begin{cases} X^i & \text{for } i < n \\ 0 & \text{for } i > n \end{cases}$$

and Y^n is set so that

$$\int_{U^n}^{\infty} G_{n+1}(Q)\,dQ = \varepsilon$$

with $0 < Y^n \leq X^n$. Because of the simplicity of this solution, an explicit nonlinear optimization algorithm is not needed to solve the subproblem, thus reducing the computation required to solve the problem. (However, calculation of the equivalent load duration curves can be computationally burdensome.)

The subproblem also gives shadow price information on the plant capacities. Define π and λ^i as the dual multipliers associated with the unserved energy constraint (4.7) and with the capacity constraint in (4.8), respectively. These multipliers must be nonnegative because the constraints are inequalities. Assuming that the subproblem is feasible and that the degenerate cases $X^i = 0$ or $Y^n = X^n$ do not arise, the Kuhn-Tucker condition for the problem give the following expressions for these shadow prices

$$\lambda^i = - \sum_{j=1}^{I} F^j p_j \frac{\partial}{\partial Y^I} \int_{U^{j-1}}^{U^j} G_j(Q;Y^1,\ldots,Y^{j-1})\,dQ$$

$$- \pi \frac{\partial}{\partial Y^I} \int_{U^I}^{\infty} G_{I+1}(Q;Y^1,\ldots,Y^I)\,dQ \quad \text{for } i < n \qquad (4.10)$$

$$\lambda^i = 0 \quad \text{for} \quad i \geq n$$

and $\pi = F^n$. The multiplier π is just the marginal cost of decreasing the unserved energy ε, which, intuitively, is the cost of operating the marginal plant. The multipliers λ^i measure the benefit of increasing the capacity X^i.

Thus, as was true for the linear and nonlinear programming models discussed previously, the operating subproblems can be solved without using an explicit optimization algorithm. The dual solution can be obtained from the Kuhn-Tucker conditions. The computational work involved in solving the subproblem arises from the probabilistic simulation recursion (4.1) and from computing the derivative terms in (4.10). The use of probabilistic simulation in the subproblems can be integrated into a capacity planning model through the use of a decomposition principle. The next section discusses the application of generalized Bender decomposition.

C. The Capacity Expansion Planning Model and Its Solution by Generalized Benders' Decomposition

Consider the structure of the capacity expansion planning model in the probabilistic case. There is an operating subproblem, described in the previous section, for each period t in the planning horizon, which can be written in the following vector form

$$\text{minimize} \quad EF_t(\underline{Y}_t) \tag{4.11}$$

$$\text{subject to} \quad EG_t(\underline{Y}_t) \leq \varepsilon_t \tag{4.12}$$

$$0 \leq \underline{Y}_t \leq \delta_t \underline{X} \tag{4.13}$$

where, as before, $\underline{X}$ is the vector of plant capacities X_{jv},

$\underline{Y}_t$ is the vector of plant utilization levels, Y^{it}, in period t, and

ε_t is the desired reliability level.

Then the objective function (4.6) for the subproblem in period t is given by the function $EF_t(\underline{Y}_t)$, the capacity constraints (4.8) are represented by (4.13), the expected unserved energy (4.7) is represented by the function $EG_t(\underline{Y}_t)$, and the matrix δ_t sorts the vector of plant capacities into merit order.

The problem of finding a minimum cost capacity expansion plan can be modeled as follows

$$\text{minimize} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t) \tag{4.14}$$

$$\text{subject to} \quad EG_t(\underline{Y}_t) \leq \varepsilon_t \quad t = 1,\ldots,T \tag{4.15}$$

$$0 \leq \underline{Y}_t \leq \delta_t \underline{X} \tag{4.16}$$

where $\underline{C}$ is the vector of plant capacity costs. (As before, for simplicity of notation, it has been assumed that all elements of the vector $\underline{X}$ are decision variables.) Written out in full, this problem takes the form

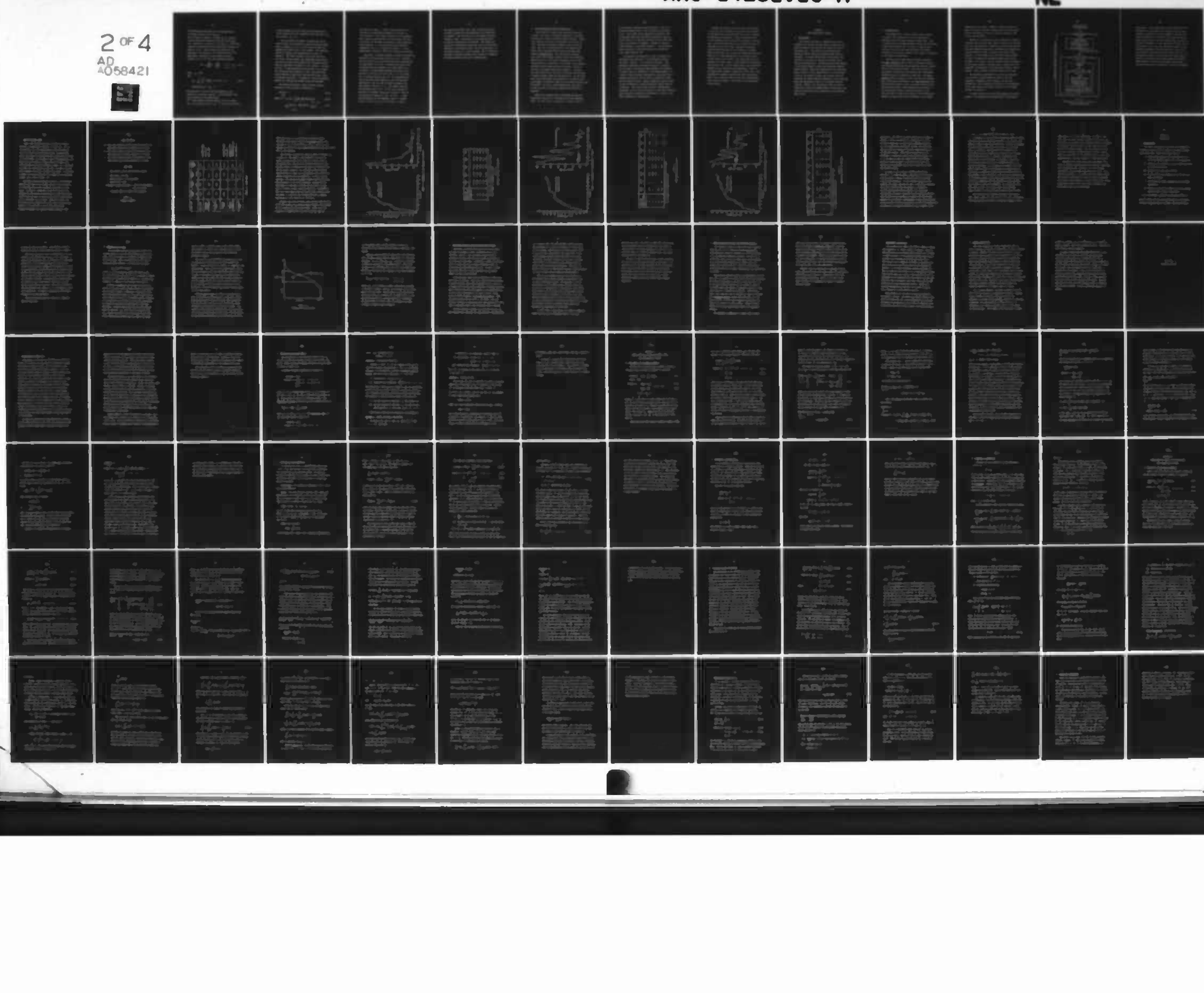
$$\text{minimize} \quad \sum_{j=1}^{J} \sum_{v=1}^{T} C_{jv} X_{jv} + \sum_{t=1}^{T} \sum_{i=1}^{I_t} F^{it} p_i \int_{U^{i-1,t}}^{U^{it}} G_{it}(Q)\,dQ \tag{4.14}$$

$$\text{subject to} \quad \int_{U^{I_t,t}}^{\infty} G_{I_t+1,t}(Q)\,dQ \leq \varepsilon_t \tag{4.15}$$

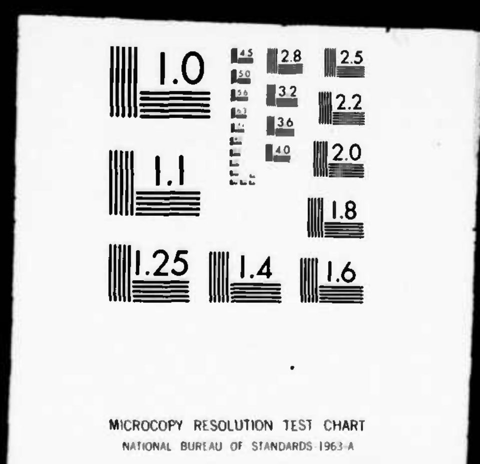
$$0 \leq Y^{it} \leq \sum_{j=1}^{J} \sum_{v=-\nu}^{t} \delta^{it}_{jv} X_{jv} \quad t=1,\ldots,T \tag{4.16}$$

AD-A058 421 MASSACHUSETTS INST OF TECH CAMBRIDGE OPERATIONS RESE--ETC F/G 12/2
DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY --ETC(U)
AUG 78 J A BLOOM DAAG29-76-C-0064
UNCLASSIFIED TR-154 ARO-14261.10-M NL

2 OF 4
AD
A058421



AD
A058421


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The Kuhn-Tucker conditions for this problem define an optimal solution to the dual, as will be shown in Chapter 8. Define $\underline{\lambda}_t$ as a vector of dual multipliers associated with the set of capacity constraints in period t (4.16) and π_t as the dual multiplier associated with the reliability constraint in period t (4.15). Then, assuming the problem is not degenerate, as discussed in the previous section, the Kuhn-Tucker conditions for the capacity planning problem (4.14)-(4.16) give the following conditions

$$\lambda^{it} + \pi_t \frac{\partial EG_t}{\partial Y^{it}} = - \frac{\partial EF_t}{\partial Y^{it}} \qquad \begin{array}{l} i = 1,\ldots,I_t \\ t = 1,\ldots,T \end{array} \qquad (4.17)$$

where $\pi_t = F^{nt}$

and

$$C_{jv} - \sum_{t=1}^{T} \sum_{i=1}^{I_t} \lambda^{it} \delta_{jv}^{it} \geq 0 \quad \text{for all } j,v \qquad (4.18)$$

with equality if $X_{jv} > 0$.

The equations (4.17) are just the equations (4.10) derived in the previous section. As noted before, they can be solved for the shadow prices λ^{it}. These shadow prices are used in (4.18) to "price out" the capacity variables X_{jv}. If the cost of building a plant exceeds the benefits derived from operating it ((4.18) holds with strict inequality) then the plant will not be built ($X_{jv} = 0$).

Just as generalized Benders' decomposition can be used to solve the deterministic capacity planning problem given in Chapter 3, so too it can be used to solve the probabilistic model given in this chapter. The technical details of applying the decomposition are given in Chapter 8. The master problem in this case also turns out to be a linear program; the subproblems are nonliear, but they are solved using probabilistic simulation and an explicit optimization procedure is not required. It turns out that a major computational difficulty is to compute the derivatives which appear in (4.10). This difficulty arises because the equivalent load duration curves $G_i(Q)$ are defined recursively. In Chapter 9, several alternative methods for computing the shadow prices are discussed.

The generalized Benders' master problem, derived according to the discussion of Chapter 8 is

$$\underset{Z,\underline{X}}{\text{minimize}} \quad Z \tag{4.19}$$

$$\text{subject to} \quad Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [EF_t(\underline{Y}_t^k) + \underline{\lambda}_t^k \delta_t(\underline{X}^k - \underline{X})] \qquad k=1,\ldots,K \tag{4.20}$$

$$\sum_{t\epsilon\Gamma_k} [EG_t(\underline{Y}_t) + \underline{\mu}_t^k \delta_t(\underline{X}^k - \underline{X})] \leq \sum_{t\epsilon\Gamma_k} \varepsilon_t \tag{4.21}$$

$$\underline{X} \geq 0$$

As before, the index k refers to trial solutions of the master and subproblems generated at iteration k. The constraints (4.21) insure feasibility in the subproblems; they are generated when a trial solution $\underline{X}^k$ does not satisfy the reliability constraint in the subproblem. The multipliers μ_t are generated by the infeasible subproblems, and Γ_k is the set of indices t of the subproblems in which the $k^{\underline{th}}$ trial solution is infeasible.

Once again, the master problem is solved by successively generating the constraints (4.20) and (4.21). Starting with an initial trial solution for the plant capacities $\underline{X}^1$, the subproblems (4.6)-(4.8) are solved for each period, to generate the first Benders' cuts (4.20) and (4.21). In general, a relaxed master problem consisting of constraints (4.20) and (4.21) with $k=1,\ldots,\ell-1$ is solved for a new trial capacity plan $\underline{X}^\ell$. The subproblems are then solved with these capacities, and the associated shadow prices $\underline{\lambda}_t^\ell$ and $\underline{\mu}_t^\ell$ are used to generate the next Benders' cuts with $k = \ell$. The shadow prices $\underline{\lambda}^\ell$ are determined from (4.17). If the trial solution X^ℓ produces an infeasible subproblem in period t (that is, the reliability constraint (4.12) is viola)), then the multipliers $\underline{\mu}_t^\ell$ are computed, using a procedure similar to (4.17), as will be discussed in Chapter 9.

The algorithm proceeds iteratively, alternating between the master problem and the subproblems until optimality is achieved. Note that the use of the complex nonlinear functions $G_i(Q)$ is confined to the subproblems where no explicit optimization is performed. The subproblems isolate the probabilistic simulation from the optimization performed in the master problem. But the optimization in the master problem is a linear program. Thus, a difficult nonlinear, stochastic program can be solved as a sequence of linear programs by the use of decomposition.

D. An Alternative Treatment of Reliability

Some authors[10], notably those discussing peak-load pricing, have taken an alternative approach to including reliability in capacity expansion planning models. Rather than constrain the level of reliability, as in (4.4) or (4.5), they have preferred to charge a cost for loss of load. This cost is representative of the economic and social costs of unserved demand and has sometimes been called a rationing cost. The rationale for charging a cost rather than setting a target is that the reliability should be set at a level where the marginal cost of providing additional reliability is just equal to the marginal rationing costs avoided by such an increase. Since, in practice, the desired reliability standard ε is often set rather arbitrarily, the use of rationing costs provides a logical economic reason for determining it. On the other hand, actual measurement and estimation of rationing costs is difficult (an important study on this subject has been made by Telson[8]), so that in actual applications, they may be no less arbitrary than direct estimates for reliability standards.

There is an important dual relationship between the reliability level and the rationing cost in the following sense - for any level of reliability ε, optimal solution of the capacity expansion planning problem (4.11)-(4.13) determines shadow prices on the reliability constraints (4.13). This shadow price represents the marginal cost of providing additional reliability and gives a ceiling on the value of the marginal rationing cost for which this level of reliability is adequate. Parametric analysis could determine a relationship between ε and the shadow price to provide a trade-off function for reliability level versus marginal cost.

An equivalent alternative is to put a multiplier on constraint (4.13) and add it to the objective function (in effect, letting ε be a decision variable). This multiplier represents the marginal rationing cost, and for any value, the capacity planning model will determine an optimal level of reliability. The problem becomes, in a sense, a multi-criterion optimization, since the cost of supplying electricity and the level of reliability would now be optimized jointly. Again, parametric analysis could be used to determine a trade-off frontier between cost and reliability. This approach may also have computational advantages, since it may be difficult to find trial solutions which satisfy the reliability constraints, at least in the early iterations of the algorithm.

Perhaps the essence of the rationing cost approach is the idea that reliability of service is an attribute of the demand for electricity and that as such it should be economically determined. Since a substantial part of the rationing costs are borne by the buyers of electricity, reliability should be set at a level for which the customers are willing to pay. However, the capacity planning model discussed in this chapter is a supply model dealing only with the costs of supplying electricity. Therefore, it takes demand and its attributes as given. In Part Three of this thesis, this supply model will be embedded in a larger model in which demand for electricity varies with price. In this larger model, the reliability standard can be regarded as price-dependent, linked to the supply model through a constraint of the (4.13), thus providing the economic rationale for these constraints. Hence, this integrated model resolves the two approaches into a single treatment. This topic is discussed again in Part Three.

CHAPTER 5

SOME COMPUTATIONAL RESULTS

A. Introduction

This chapter discusses an implementation of the probabilistic capacity planning model, presented in the previous chapter, and presents the results of some experimental runs. The algorithm was implemented by modifying the MIT Generation Expansion Model (GEM) of Schweppe[1]. The major modifications required were the addition of a routine to calculate the shadow prices within the probabilistic simulation and the modification of the linear program to solve the generalized Benders' master problem. Three test problems were run using data based (loosely) on the characteristics of a New England utility. The results of these runs indicate that the algorithm can indeed produce a sequence of trial solutions which get successively closer to an optimal feasible solution. However, the runs were not able to achieve feasibility within the allowed number of iterations, and there are indications that convergence of the algorithm may be show.

B. Implementation

The probabilistic capacity planning model proposed in Chapter 4 was implemented for testing purposes by modifying the MIT Generation Expansion Model (GEM). GEM is a detailed utility planning model intended for production (rather than research) use. As such, it contains many facilities which have not been included in the models discussed in this thesis; however, it is similar in structure to the models discussed here.

GEM consists of three major, integrated submodels. The plant evaluation model is used to determine feasible plant designs which meet the environmental quality standards set for various types of sites. This model screens out unacceptable alternatives and evaluates plant performance characteristics for acceptable alternatives. The plant expansion model determines a least cost capacity expansion plan, using linear programming. The plant operation model determines the operating costs of the plants built by the plant expansion model, using probabilistic simulation. The relationship between the expansion model and the operation model in GEM is very similar to the relationship between the master problem and subproblems in the models presented in this thesis. However, in GEM, the operation model is used to calculate a capacity factor for each plant, and this information is used to calculate operating costs in the expansion model.

Because of its similarity to the probabilistic capacity planning model described here, GEM was chosen as the starting point for implementation. Specifically, many of the routines and data structures required for implementing the model were already available in GEM. Two major modifications were required. In the probabilistic simulation program SYSGEN[2], a routine to compute the subproblem shadow prices had to be added. This routine uses the formulas derived in Chapter 9. In the linear program, a different type of matrix had to be generated, corresponding to the generalized Benders' master problem. Also, the plant evaluation model was not used.

The algorithm implemented for solving the subproblem is an earlier, slightly different version of the one discussed in Chapters 8 and 9; however, for the test problems run, this difference is believed to be unimportant. A flow chart showing the overall structure of the program is given in Figure 5.1.

In order to test the routine which computes the shadow prices, a test problem was designed based on the Binomial

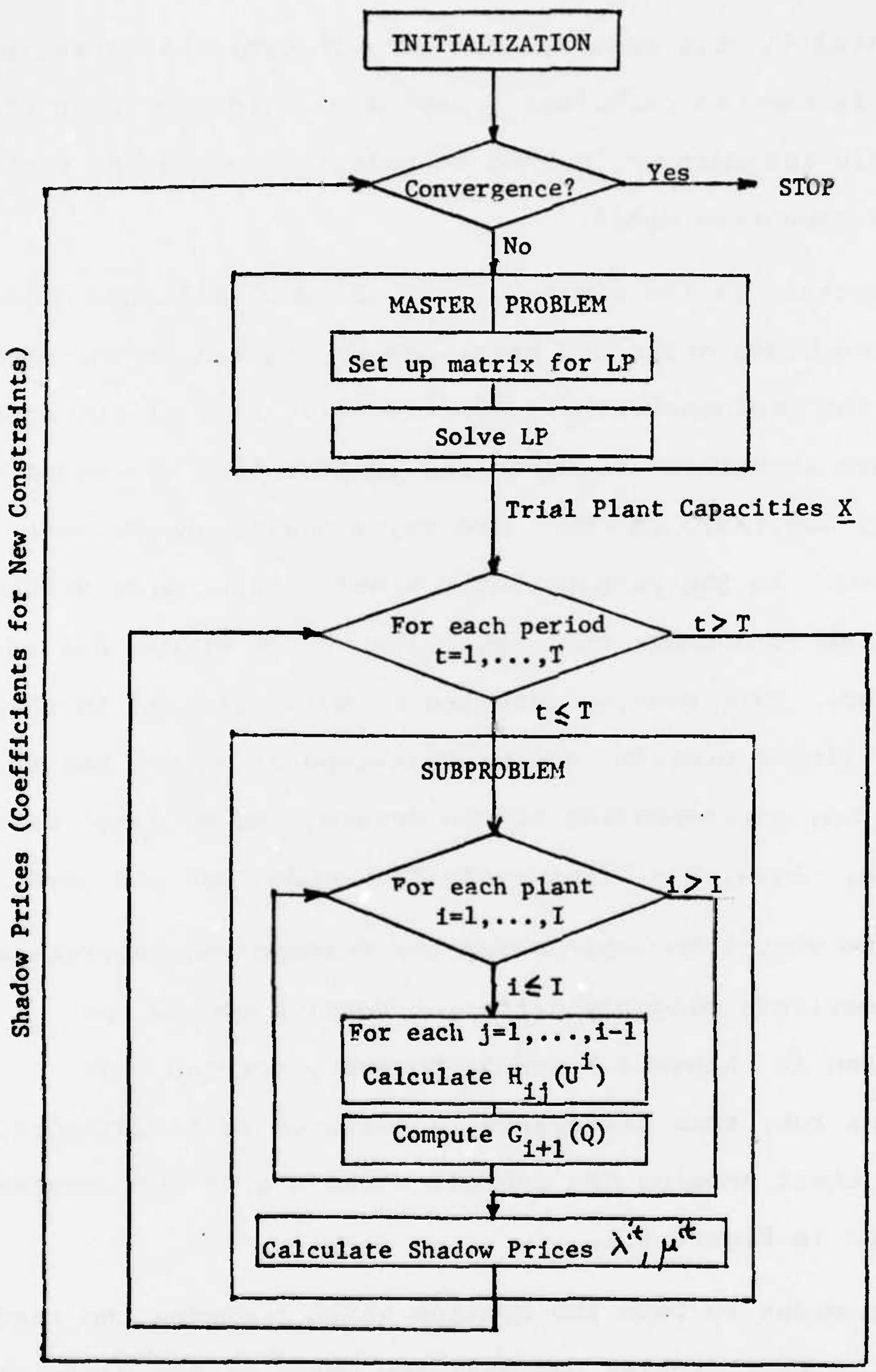


Figure 5.1

The Algorithm for Solving the Probabilistic Capacity Planning Problem

distribution. This distribution has the property that the convolution of two Binomial distributions with the same probability parameter is again a Binomial, and the order of this distribution is the sum of the orders of the two which were convolved. The test problem used a cumulative Binomial distribution for the load duration curve, and all plants were of unit size with availability equal to the probability parameter of the distribution. Thus, all of the equivalent load duration curves and the functions H_{ij} (discussed in Chapter 9) should also have been Binomial. Therefore the answers given by the computation could be checked against a table of the Binomial distribution. This problem provided a convenient test case for debugging.

C. Results of Test Runs

Three test problems were run. The data for these problems is summarized in Tables 5.1 and 5.2. In all the problems, the same load duration curve was used in the first year of the planning horizon. In subsequent years, this curve is scaled up by the growth rate specified for the problem. In each year of the planning horizon, the model can build plants of three alternative types, the characteristics of which are given in Table 5.2. Although the model can build plants of any size, the data are given for a standard size plant, listed in the table. In addition to these new plants, there are five committed and existing plants, also listed in Table 5.2.

The short planning horizons used tend to discriminate against high capital cost plants, since the full benefits of installing these plants cannot be recovered in a few years. In order to compensate for this effect, it was assumed that the system configuration and load characteristics of the last year of the planning horizon would prevail in all subsequent years, forever. Thus, each plant will be replicated at the end of its life, and all plants will operate in the same way as in the last year of the horizon throughout the infinite extension period. The

TEST PROBLEMS

A. Four year planning horizon without extension costs

Load growth rate 40%/year (doubles in 2 years)

B. Four year planning horizon with extension costs

Load growth rate 40%/year (doubles in 2 years)

C. Nine year planning horizon with extension costs

Load growth rate 8%/year (doubles in 9 years)

LOAD DATA

For initial year of the planning horizon

Peak Load - 2100 MW

Energy Demand - 11,275,000 MWH

Load Factor - 60%

Unserved Energy Constraint $\leq$ 2.8% of Energy Demand (for all years)

Discount Rate is 10.8% in all problems

TABLE 5.1

PROBLEM SUMMARY

PLANT TYPE	STANDARD SIZE MW	CAPITAL COST $/KW	OPERATING COST $/MWH	FORCED OUTAGE RATE %
LWR, BASE	1000	970	6.12	31.5
COMBINED CYCLE, INTERMEDIATE	700	500	17.9	22.7
GAS TURBINE PEAK	300	240	49.0	13.8
NUCLEAR, BASE	1000	-	6.22	31.5
COMBINED CYCLE, INTERMEDIATE	700	-	19.8	22.7
GAS TURBINE PEAK	300	-	49.1	13.8

ALTERNATIVES, AVAILABLE EACH YEAR

COMMITTED AND EXISTING.

NUMBER GIVEN:
NUCLEAR, 1
COMBINED CYCLE, 2
GAS TURBINE, 2

TABLE 5.2 PLANT DATA

discounted costs of replicating and operating plants throughout this period, called the extension costs, were included in two of the problems, B and C. (The nine year horizon of Problem C is the longest permitted by the data arrays used in the program.)

The results of these test runs are shown in Figures 5.2, 5.3 and 5.4. The graphs in these figures show how the linear program optimal value, which is a lower bound on the optimal cost, and the total unserved energy over the horizon, which is a measure of the degree of infeasibility of the current solution, vary with the iteration number. (Again,because of array size limitations, only nineteen or twenty iterations could be performed.) As these figures indicate, the iteration procedure moved toward feasibility and optimality in all of the problems. However, within the number of iterations allowed, the algorithm did not find an optimal or even a feasible solution. Furthermore, the trial plant capacities generated at each iteration did not appear to stabilize but instead continued to change significantly from iteration to iteration.

Because the trial solutions never attained feasibility, it was not possible to establish upper bounds on the optimal cost to compare with the lower bounds. Hence it is not possible to determine how close the algorithm came to

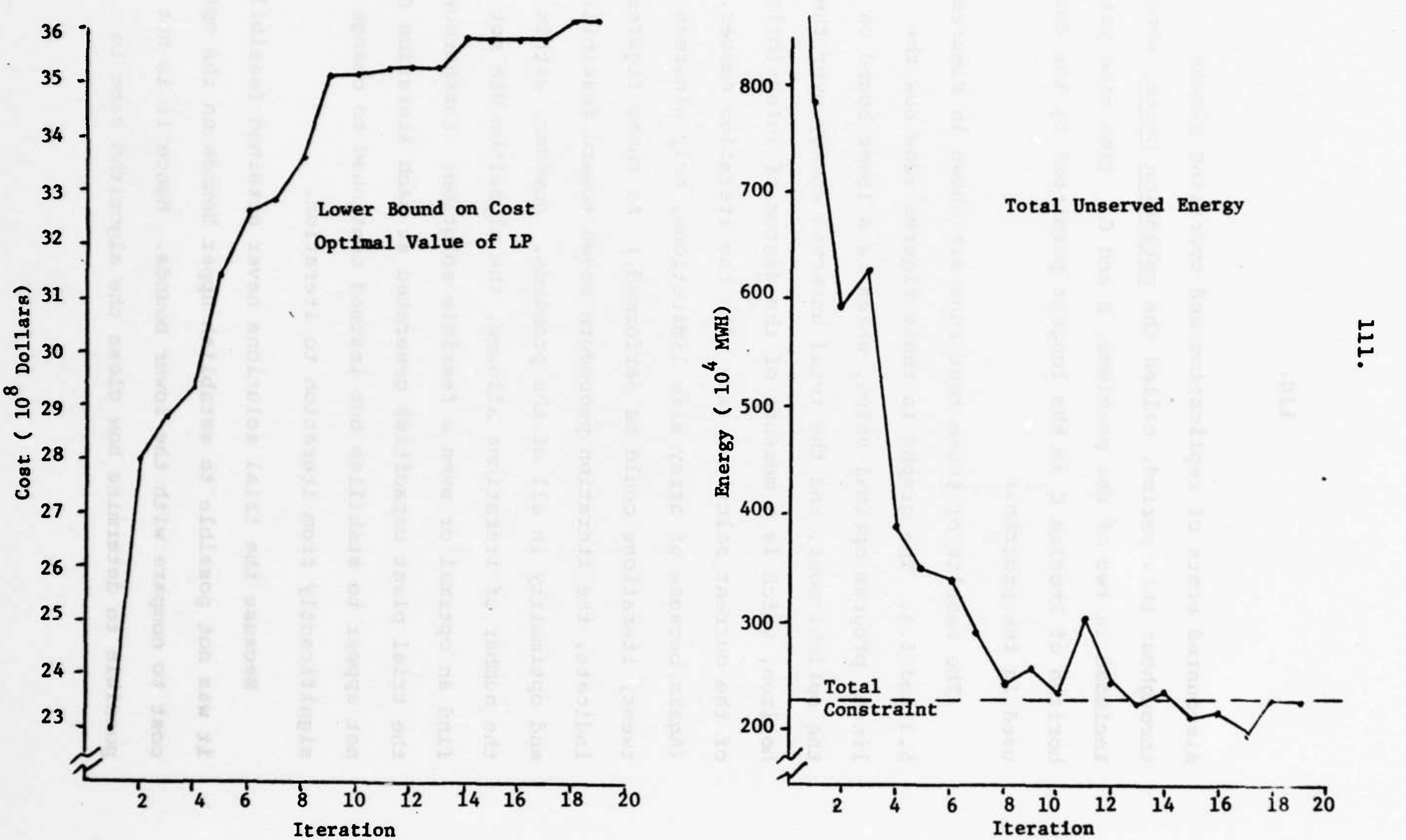


Figure 5.2 Convergence of the Algorithm in Problem A

PLANT TYPE VINTAGE YEAR*	INTR. 3	PEAK 1	PEAK 2	PEAK 3	PEAK 4
ITERATION NO.					
15	0.524	2.099	0.275	2.063	5.053
16	0.336	1.467	0.359	3.398	4.804
17	0.109	1.919	0.632	3.099	4.858
18	0.156	0	1.964	3.655	5.205
19	0.560	0.307	1.998	1.916	5.630

*No other plant alternatives were chosen by the model.

Units: Number of standard-size plants.

Table 5.3

Selected Trial Plans from Problem A

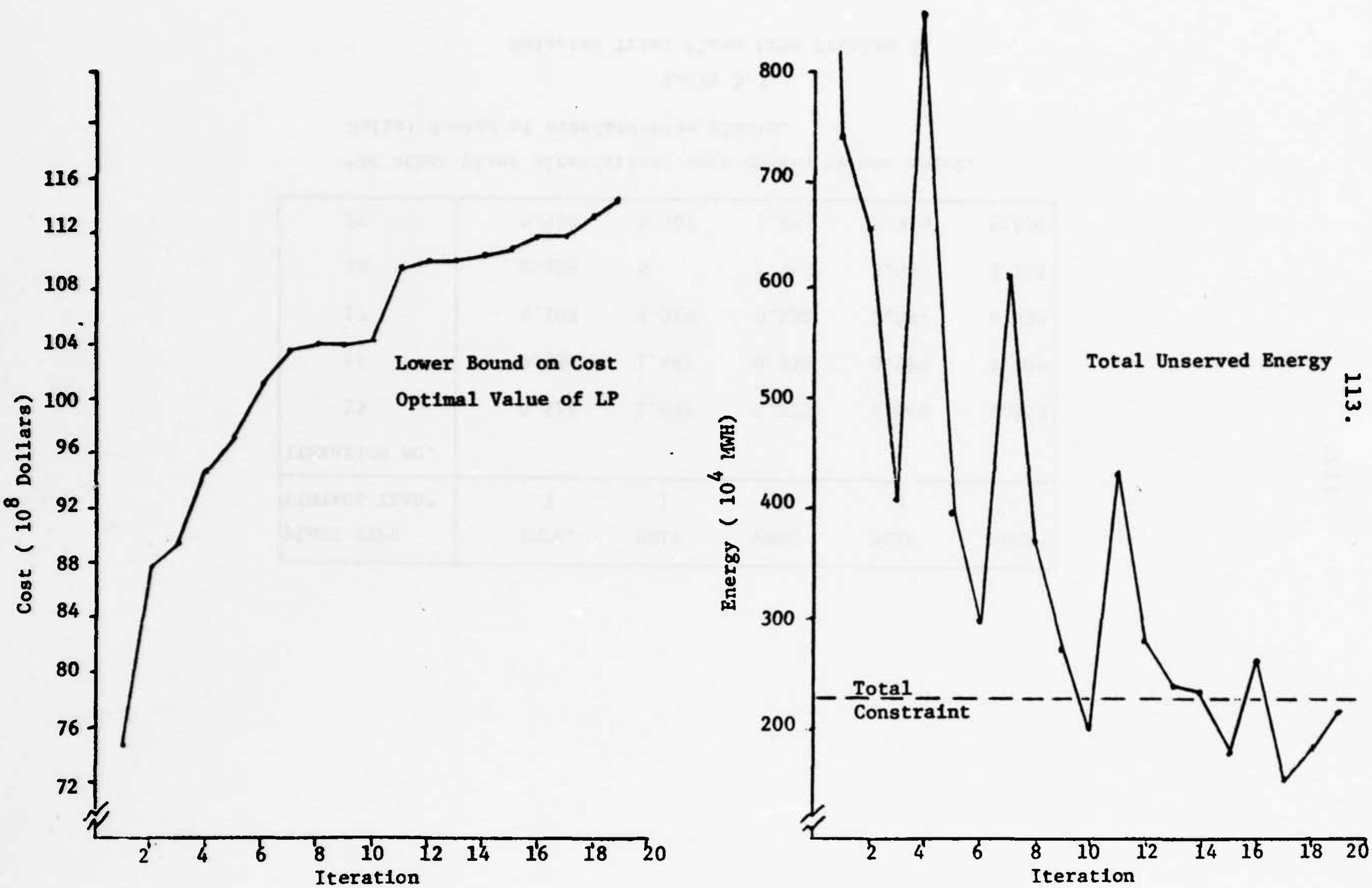


Figure 5.3 Convergence of the Algorithm in Problem B

PLANT TYPE VINTAGE YEAR*	BASE 1	BASE 2	BASE 3	BASE 4	INTR. 1	INTR. 2	INTR. 3	INTR. 4
ITERATION NO.								
15	1.02	0	0.50	0	0.25	0.42	1.69	0.75
16	0	0	0.98	0.83	0.02	2.15	0	0.59
17	0	0	0.32	0.51	0.74	1.78	0.95	0.76
18	0.06	0	0	0	1.98	0	1.87	1.67
19	0.58	0.77	0	0.37	0.51	0	0	2.81

*No other plant alternatives were chosen by the model

Units: Number of standard-size plants.

Table 5.4

Selected Trial Plans from Problem B

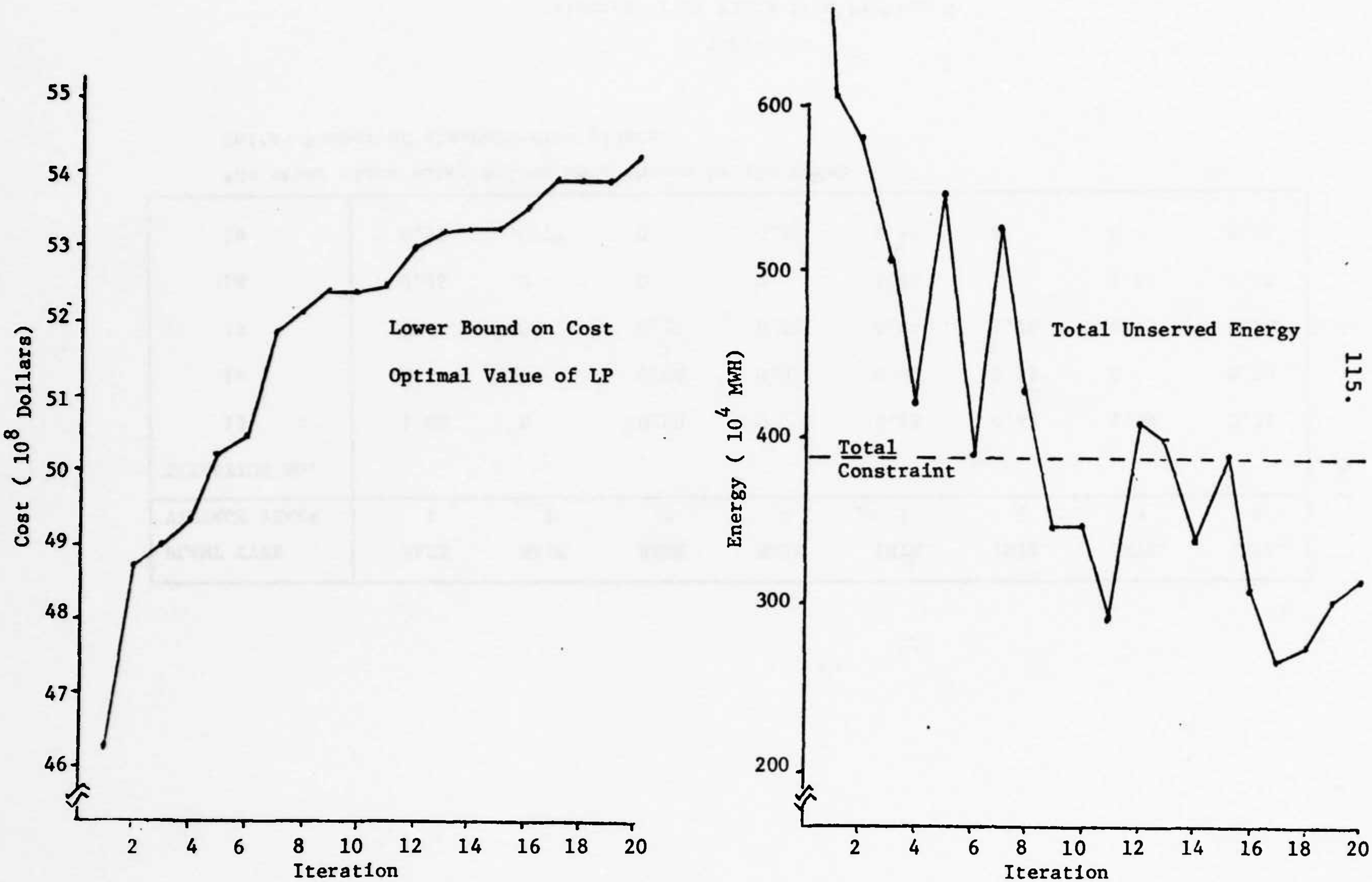


Figure 5.4 Convergence of the Algorithm in Problem C

PLANT TYPE VINTAGE YEAR*	INTR. 1	INTR. 2	INTR. 3	INTR. 4	INTR. 5	INTR. 6	INTR. 7	INTR. 8	INTR. 9
ITERATION NO.									
16	0	0.21	0	0.70	0.68	0	0.10	0.43	0
17	0.39	0.66	0	0	0.09	0	1.03	0	0
18	0	1.08	0.14	0	0.44	0	0	0.40	0.10
19	0.70	0	0.22	0	0.32	0	0.29	0	0.70
20	0.06	0.03	0.81	0	0.22	0	0.58	0.41	0.16

*No other plant alternatives were chosen by the model.

Units: Number of standard-size plants.

Table 5.5

Selected Trial Plans from Problem C

optimality. The failure to attain feasibility is easy to explain. First, the reliability constraint is a hard, nonlinear constraint so that finding feasible solutions is, not unexpectedly, difficult. Second, the algorithm used to solve this problem uses outer approximations of the feasible region. These outer approximations are polyhedral in shape and contain the feasible region. It is to be expected that the linear program will always generate a solution which is an extreme point of this approximating polyhedron, which will always lie outside the feasible region.

In order to try to find a feasible solution in Problem A, the unserved energy constraints were relaxed slightly after a number of feasibility constraints had already been generated. In this relaxed version of the problem, the master problem was able to generate a feasible solution which satisfied the optimality conditions. However, because the master problem still contained feasibility constraints generated by the unrelaxed version, this solution is not necessarily optimal for the relaxed version of the problem. Nevertheless, a more sophisticated version of this relaxation procedure, in which the feasibility constraints in the master problem are also relaxed, could be used to find upper bounds for the optimal cost.

In Problem A, which did not include the extension costs, the trial solutions consisted entirely of low capital cost peaking plants in the early iterations. In later iterations, one intermediate plant began to show up. On the other hand, in Problem B, the trial solutions consisted of intermediate plants in the early iterations and of intermediate and base plants in the later iterations. No peaking plants were proposed. In Problem C, the trial solutions consisted entirely of intermediate plants in all the iterations. It is conjectured, however, that the master problem tends to avoid plants with higher capital costs until a number of Benders' cuts have been generated. The economy of the high capital cost plants is not expressed in the master problem until a good approximation to the cost function is obtained, and this requires a number of cost constraints to have been generated. Since Problem C has more time periods, it is likely that more cuts were required before a good enough approximation was obtained.

The algorithm also tended to generate trial plans with a small number of large plants in the early iterations, with more uniform distribution of capacity expansion over time showing up in later iterations. In fact, in some test runs, the master problem proposed a plant that was too big for the probabilistic simulation routine to handle; a constraint limiting the size of the plants had to be added to the master problem. The non-uniformity of all the early iterations is probably due to having a small number of constraints in the master problem, so that the approximation to the cost function is very non-uniform.

The experience gained from these experimental runs tends to indicate that the generalized Benders' algorithm for this problem works reasonably, but that further improvements could be made with additional study. Particularly important is the need to generate upper bounds (and feasible solutions) in order to be able to measure the distance from optimality. Further experiments, using larger data arrays, should also help to determine the convergence rate for the algorithm.

CHAPTER 6

EXTENSIONS

A. Introduction

The models presented so far have used a number of simplifying assumptions, in order not to obscure their basic structure. These assumptions include:

i) Only thermal plants, and not hydroelectric or other non-thermal plants, have been considered.

ii) Capacity and operating costs have been represented as linear functions.

iii) Plants of any size can be built.

iv) Plant location and transmission costs have not been considered.

v) Environmental quality standards have not been considered.

vi) Multiple valve-point plants and use of spinning reserves have not been considered.

vii) Maintenance planning has not been included.

Many of these features could be included in the models without disrupting their structure, and decomposition methods can still be applied. This chapter discusses briefly how some of these extensions could be handled.

There are two basic types of modifications that can be made to the models which preserve their structure - those which affect the capacity variables, and therefore become part of the master problem, and those which affect the operating variables, and therefore become part of the subproblem. Furthermore, since the decomposition approach derives much of its utility from the fact that special algorithms can be used to solve both the subproblem and the master problem, the modifications should either be compatible with these special algorithms or have special algorithms of their own. The special algorithms for the subproblem all depend on the optimality of merit order operation. The special algorithm for the master problem is linear programming.

The following sections each discuss a different type of extension.

B. Hydroelectric Plants

The difference between hydroelectric plants and thermal plants is that the total energy generated by a hydro plant is limited by the amount of water stored behind the dam. Thus, the simplest model of hydroelectric plants includes a constraint of the form

$$\int_{\tau \epsilon I} Y_{hv}(\tau)\, d\tau \leq H_{hv}(I)$$

where h is the plant-type index for hydro plants, I is a time interval, and $H_{hv}(I)$ is the amount of energy available to the plant during interval I. Usually this interval is a season or some shorter time period.

A special methodology for placing hydroelectric plants in the merit order has been developed by Jacoby[1] and extended by others. Generally, hydro plants have negligible operating costs, so it would be desirable to place them at the bottom of the loading order. However, it is generally not possible to do so since they would be required to generate more energy than they have available. Jacoby's rule states that a hydroelectric plant of capacity X_{hv} and with given stored energy H_{hvt} should be placed in the merit order so that the total energy generated (the area cut out of the load duration curve) is exactly equal to the stored energy (see Figure 6.1). Then merit order operation is still the optimal policy, and solution of the subproblem proceeds as before.

This rule has been extended to loading hydro plants in probabilistic simulation by Joy and Jenkins[2] and Finger[3]. Finger also considers loading of pumped-storage hydro plants. Thus, hydro and even pumped hydro plants can be considered within the subproblems already defined. In order to calculate shadow prices for these plants (shadow prices will appear not only for power capacity, as with thermal plants, but also for energy capacity), it is necessary to add the proper constraints to the subproblem which characterize the optimal loading position of these plants.

Thermal plants under air quality constraints can also be modelled like hydro plants, if the air quality standard constrains total emissions. The constraint then has the same form as an energy constraint, since emissions are proportional to total fuel burned. The loading of these plants is complicated by the fact that their operating costs are non-negligable, but the effect of the emissions constraint is to push the plant higher

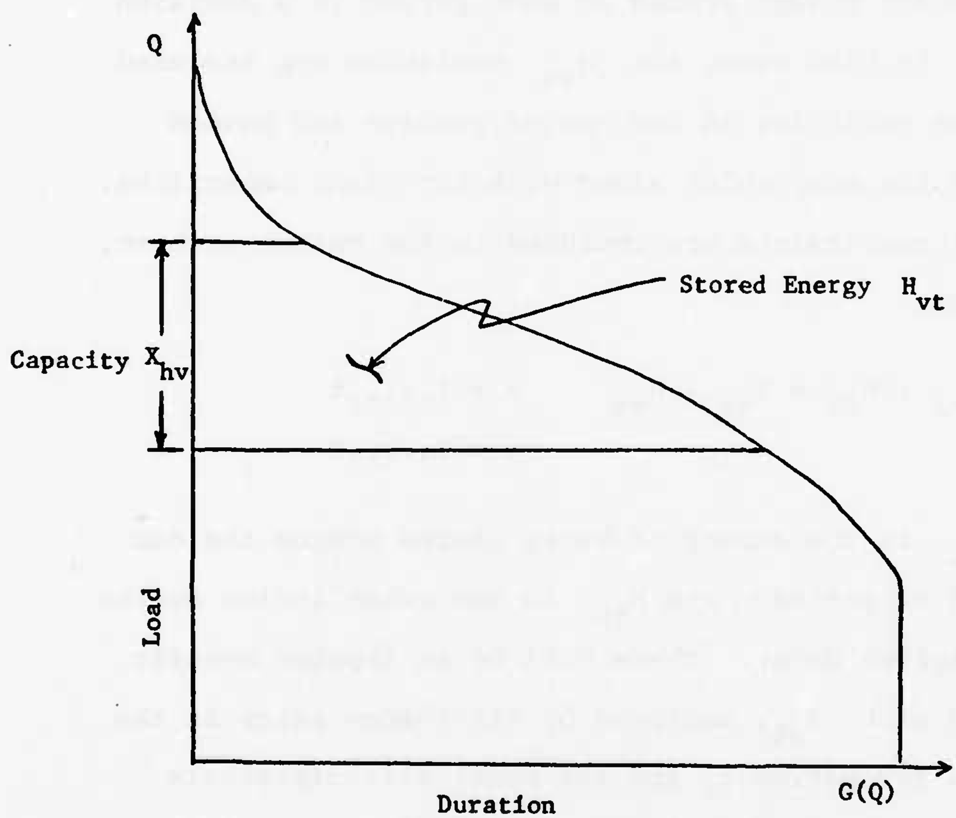


Figure 6.1

Loading a Hydroelectric Plant

in the loading order, if the constraint cannot be satisfied at its normal position (based on operating cost alone).

Benders' decomposition can also handle the situation when the hydro energy stored in each period is a decision variable. In this case, the H_{vt} variables are included as decision variables in the master problem and passed as data to the subproblem along with the plant capacities. Additional constraints are included in the master problem, of the form

$$S_{v,t-1} + W_{vt} = S_{vt} + H_{vt} \qquad \begin{array}{l} v = 1,\ldots,t \\ t = 1,\ldots,T \end{array}$$

where S_{vt} is the amount of water stored behind the dam at the end of period t, and W_{vt} is the water inflow during period t (given data). There will be an imputed benefit associated with H_{vt}, measured by its shadow price in the subproblem for period t, and the model will distribute the hydro energy available among the different periods in order to optimize these benefits.

C. Nonlinear Capacity Costs and Fixed-Size Plants

The models discussed previously have assumed that the cost of building a plant is proportional to its capacity. However, real generating plants exhibit economies of scale: the cost per kilowatt decreases as the plant size increases. Furthermore, plants are usually built in fixed unit sizes because components such as generators are manufactured that way. In order to deal with these features, integer variables can be introduced into the master problem.

Cost functions which exhibit economies of scale are nonlinear and concave. They can be approximated piecewise linearly; however, when such an approximation is used in a cost-minimizing optimization problem, an integer variable must be associated with each segment of the approximation to insure that the segments are used in proper order. Thus the problem becomes a mixed integer program. When fixed-size plants are included in the model, a binary decision variable is associated with each alternative plant size to indicate whether that particular size is chosen.

Integer variables present no additional problems to the decomposition approach; in fact, Benders' decomposition is a standard tool for solving mixed integer programs. The decomposition is used to separate the integer part of the problem, which becomes the master problem, from the linear part, which becomes the subproblem. With proper design, the operating subproblems will remain unchanged. The master problem will become an integer program which can be solved by the standard cutting plane or branch-and-bound techniques. (The paper by Noonan and Giglio[4] in fact, applies Benders' decomposition to solve a mixed integer capacity planning model).

The use of a concave capacity cost function in the probabilistic capacity planning model should lead to some interesting results. The economies of scale should tend to favor large plants; however, there is a diseconomy of reliability in large plants. Since fewer large plants will be built, there will be less diversification and lower reliability (this occurs even when expected unserved energy is used to measure reliability). It will be interesting to determine the plant size at which the economy of scale in building large plants just balances the diseconomy of reliability.

In Part Three of this thesis, the capacity planning model will be used to calculate marginal costs for peak-load pricing. It should be noted that calculation of such marginal costs in a model which uses concave cost functions or integer variables is very much more difficult than in a linear or convex model. This area deserves further investigation.

Finally, it should be noted that integer variables can also be used for other purposes in this model. An important use is to model available siting alternatives. Once a site has been used, it cannot be used again. Furthermore, there may be restrictions on the type of plants which can be built on a given site. These additional constraints can be modeled using integer variables.

D. Multi-Valve-Point Plants and Spinning Reserves

The operating range of a generating plant actually consists of a set of intervals between valve points. Often operating costs are different in different intervals and sometimes different intervals can fail independently. Hence a multiple valve point representation can be used to model nonlinear operating costs and probability distributions for available capacity more complicated than the two point distribution used previously, in Chapter 4.

When plants' operating ranges are represented in this way, the valve points of a given plant may occupy different, non-adjacent positions in the merit order, and thus the merit order operating scheme may no longer be optimal (an upper valve point cannot be operated unless the lower valve points are already in operation). Furthermore, merit order operation may be violated in order to provide for a spinning reserve (plants kept operating at a low level in order to allow rapid response to increases in load or to failures of other plants). Thus, in these models, solving the subproblems becomes less straightforward.

The operating subproblems in models which include multi-valve-point plants and spinning reserve requirements have implicit precedence constraints which may require certain plant segments to operate before others. These precedence constraints give the operating problem a combinatorial aspect. Because of the discrete nature of these constraints, it may not be possible to define shadow prices for them, and that disrupts the relationship between the master and subproblems.

However, work is progressing on operating models which include multiple valve point plants and precedence constraints, particularly on incorporating them into probabilistic simulation. Hence, it is likely that these features will eventually be usable in capacity planning models.

E. Maintenance Scheduling

The probabilistic model presented in Chapter 4 deals with unplanned, or forced, outages of plants. However, much maintenance is performed during planned outages, and it would be useful if the capacity planning model could schedule these outages to minimize the added costs. This feature can be added to the model by dividing each year of the planning horizon into several maintenance periods. In the master problem, integer variables are used to indicate which plants will be assigned to maintenance during each such period. These plants will not be presented to the subproblem for that period, but the subproblems remain unchanged except that there must be one subproblem for each maintenance period. Again, the Benders' decomposition approach can be used to solve this problem as a mixed integer program.

Some thought must be given to logically defining maintenance periods. The definition should separate high demand periods from low, since maintenance is generally performed in low demand periods, while all capacity that can be made available is used during high demand periods.

F. Demand Uncertainty

The model proposed in Chapter 4 considers the uncertain aspects of supplying electricity, the random failure of plants. However, the demand for electricity is also uncertain, and the impact of this uncertainty is at least as great as that of the supply uncertainty on planning. There are two types of uncertainty in demand, long-term and short-term. The long-term uncertainty is primarily concerned with the gross parameters of demand, such as yearly peak, yearly energy, and load factor, several years ahead. The short-term uncertainty is primarily concerned with the fine variations which will occur hourly or daily over the next year or so. Of course, both types of uncertainty are really different aspects of the same thing, but different methods can be used to deal with them.

For planning purposes, long-term uncertainty can be dealt with by considering alternative scenarios. Usually, a small number of scenarios will be sufficient to represent the bounds of the expected variation in long-term forecasts. For each scenario, a demand forecast is developed, and a capacity planning study is done based on that forecast. Then, based on the likelihood of each scenario occurring, one or a combination of the capacity plans is chosen. This choice procedure could be formalized as a decision analysis problem.

In order to deal with short-term uncertainty, it is necessary to find a way to represent the short-term random characteristics of the demand profile. To a certain extent, the load duration curve already represents some of these characteristics; however, though it is often used as one, the load duration curve is not exactly a probability distribution. Thus, it would be useful to be able to include the random characteristics of demand within the load duration curve. Steps in this direction have been taken by Vardi, et al.[5], and others. Another approach, not using the load duration curve, is taken by Crew and Kleindorfer[6]. However, additional research is needed.

Part Two

Technical Issues

A. Organization of Part Two

The purpose of Part Two is to rigorously derive procedures for solving the two nonlinear programming models, presented in Chapters 3 and 4. The algorithm for the deterministic model is discussed in Chapter 7, that for the probabilistic model in Chapter 8. Some computational aspects of the probabilistic model are discussed in Chapter 9. The algorithm proposed for both models is the generalized Benders' decomposition of Geoffrion [17]. This algorithm proceeds by alternately solving a master problem to determine a trial soltuion for the optimal capacity expansion plan and then a set of subproblems to determine the cost of operating this trial system in the different years of the planning horizon. The shadow prices determined in these subproblems are used to generate new constraints in the master problem, which is then resolved to determine a new trial soltuion. One of the constraints generated approximates the nonlinear operating cost function; the other approximates the feasible region of the problem.

In both the deterministic and probabilistic versions of the model, presented in Part One, the subproblems have special structure which makes solving them fairly easy.

Application of the decomposition principle permits this special structure to be exploited in solving the larger capacity planning problems. Relatively simple computational procedures based on these special structures can be used to find optimal solutions and dual multipliers for the subproblems, which are used to generate the cost and feasibility constraints in the master problem.

The presentations of the algorithms in the following chapters both follow the same general steps, which are based on the arguments of Geoffrion. First, the generalized Benders' master problem is derived. Second, the optimal solution of the subproblem and the derivation of its dual multipliers are discussed. Also included are discussions of the convexity and duality properties of the subproblems, which justify the derivation of the master problem. Third, the situation when the subproblem is infeasible is discussed and dual multipliers for this situation are derived. Finally, the situation when the solution of the subproblem is degenerate is discussed.

These technical issues are presented in Chapter 7 for the deterministic problem and in Chapter 8 for the probabilistic model. Since the technical discussion for the probabilistic model is somewhat more involved, the deterministic model is presented first, in order to provide a simpler demonstration of the arguments. Chapter 9, the final chapter of this part, discusses the computational aspects of solving the subproblems in the probabilistic model.

The remainder of this introduction states some key results from the duality theory of nonlinear programs, which will be used extensively in the derivations of the following chapters.

B. Results from Duality Theory

The derivations in the following chapters rely on certain results from the duality theory of nonlinear programs. These results are summarized below; the proofs may be found in Lasdon [23].

A general nonlinear program can be stated as

$$\begin{array}{rlll} \text{minimize} & f(\underline{x}) & & \\ \text{subject to} & g_i(\underline{x}) \leq 0 & & i = 1,\ldots,\ell \\ & g_i(\underline{x}) = 0 & & i = \ell+1,\ldots,m \\ & \underline{x} \in S & & \end{array}$$

where $\underline{x}$ is an n-dimensional vector, S is a subset of R^n and f and g_i are real-valued functions defined on S. This problem is known as the primal problem. The Lagrangian function associated with this primal problem is

$$L(\underline{x},\underline{\lambda}) = f(\underline{x}) + \sum_{i=1}^{m} \lambda_i g_i(\underline{x})$$

with $\lambda_i \geq 0$ for $i = 1,\ldots,\ell$. The Lagrangian dual of the nonlinear program stated is to

$$\begin{array}{rl} \text{maximize} & L(\underline{\lambda}) \\ \text{subject to} & \lambda_i \geq 0 \quad \text{for} \quad i = 1,\ldots,\ell \end{array}$$

where $L(\underline{\lambda})$ = minimum $L(\underline{x},\underline{\lambda})$
subject to $\underline{x} \in S$

Theorem 1: Kuhn-Tucker Conditions

Suppose $S = R^n$. If $\underline{x}^*$ is a local minimum in the nonlinear program, the functions f and g_i are differentiable, and the problem satisfies a constraint qualification, then there exist multipliers $\underline{\lambda}^*$ such that

i) Stationarity: $\nabla_x L(\underline{x}^*,\underline{\lambda}^*) = 0$

ii) Complementary Slackness: $\lambda_i^* g_i(x^*) = 0 \quad i = 1,\ldots,\ell$ ||

The Kuhn-Tucker conditons give necessary, but not sufficient, conditions for optimality. There are a number of alternative constraint qualifications which guarantee the theorem. For purposes of this discussion either of two suffices:

i) All the constraint functions g_i are linear, or

ii) All the constraint functions g_i are convex and there exists an $\underline{x}$ such that $g_i(x) < 0 \quad i = 1,\ldots,\ell$.

Sufficient conditions for optimality are provided by

Theorem 2: Global Optimality Conditions

If there exist $\underline{x}^* \in S$ and $\underline{\lambda}^*$ (with $\lambda_i^* \geq 0$ for $i = 1,\ldots,\ell$) such that

i) Minimality: $\underline{x}^*$ minimizes $L(\underline{x},\underline{\lambda}^*)$ over S

ii) Feasibility: $g_i(\underline{x}^*) \leq 0 \quad i = 1,\ldots,\ell$

$$g_i(\underline{x}^*) = 0 \quad i = \ell+1,\ldots,m$$

iii) Complementary Slackness: $\lambda_i^* g_i(\underline{x}^*) = 0, \; i = 1,\ldots,\ell$

then $\underline{x}^*$ is optimal in the primal and $\underline{\lambda}^*$ is optimal in the dual. ||

Theorem 3: Strong Duality

The points $\underline{x}^*$ and $\underline{\lambda}^*$ satisfy the global optimality conditions if and only if $\underline{x}^*$ is feasible in the primal, $\underline{\lambda}^*$ is feasible in the dual and $f(\underline{x}^*) = L(\underline{\lambda}^*)$. ||

It is not true, however, that if $\underline{x}^*$ and $\underline{\lambda}^*$ are optimal in their respective problems then

$$f(\underline{x}^*) = L(\underline{\lambda}^*).$$

There can arise situations, called duality gaps, in which

$$f(\underline{x}^*) > L(\underline{\lambda}^*).$$

The global optimality conditions are sufficient, but not necessary, for optimality. If, however, f and g_i are differentiable and convex (affine, for $i = \ell+1,\ldots,m$), the global optimality conditions are equivalent to the

Kuhn-Tucker conditions, and are both necessary and sufficient.

In general, differential stationarity conditions, such as (i) of the Kuhn-Tucker conditions, are only necessary for local optimality. With convexity, however, they are sufficient to guarantee global optimality because a stationary point of a convex function is a global minimum.

CHAPTER 7

SOME TECHNICAL ISSUES ASSOCIATED WITH THE DETERMINISTIC MODEL

A. Derivation of the Master Problem

Recall that the deterministic capacity expansion planning model is stated in Chapter 3 as follows

$$\underset{\underline{X},\underline{U}_1,\ldots,\underline{U}_T}{\text{minimize}} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} F_t(\underline{U}_t) \tag{7.1}$$

$$\text{subject to} \quad M_t\underline{U}_t = \delta_t\underline{X} \tag{7.2}$$

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

$$N_t\underline{U}_t \geq Q_t^* \tag{7.3}$$

$$\underline{X} \geq 0 \quad \underline{U}_t \geq 0$$

where $\underline{X}$ is the vector of plant capacities to be built, and $\underline{C}$ is the vector of plant capital costs per unit of capacity. The matrix δ_t selects the capacities of the plants available in period t and arranges them in merit order. The vector $\underline{U}_t$ represents the plant loading points in period t, and Q_t^* is the peak load in that period.

Associated with this capacity planning problem, there is a set of operating subproblems, one for each time period in the planning horizon. They have the form

$$\underset{U^1,\ldots,U^I}{\text{minimize}} \quad \sum_{i=1}^{I} F^i \int_{U^{i-1}}^{U^i} G(Q)\,dQ \qquad (7.4)$$

$$\text{subject to} \quad U^i - U^{i-1} = X^i \quad i = 1,\ldots,I \qquad (7.5)$$

$$U^I \geq Q^* \qquad (7.6)$$

$$U^i \geq 0$$

(the index t indicating the time period has been suppressed for clarity). Here, U^{i-1} is the loading point for the i^{th} plant, the cumulative capacity of all the plants below it in the merit order ($U^0 = 0$). The load duration function, $G(Q)$, shows the length of time, during the period, during which the load exceeds level Q. The cost of operating plant i per unit of energy produced is represented by F^i, and the indexing scheme, the merit order, is designed so that $F^i \leq F^{i+1}$. The capacity X^i of the i^{th} plant is considered given data in the operating subproblems.

These operating subproblems are included as parts of capacity planning problem. The plant loading points U^i in the subproblem for period t form the vector $\underline{U}_t$ in the capacity planning problem, the objective function (7.4) is $F_t(\underline{U}_t)$, the loading order constraints (7.5) are represented by the matrix M_t, and the vector N_t represents the peak load constraint (7.6). The plant capacities in merit order X^i are generated by $\delta_t\underline{X}$.

The capacity planning problem (7.1) - (7.3) can be written in equivalent form as a two-stage optimization

$$\underset{\substack{\underline{X} \geq 0 \\ \underline{X} \,\varepsilon\, \Omega}}{\text{minimize}} \left\{ \underline{C}'\underline{X} + \sum_{t=1}^{T} \left| \begin{array}{ll} \underset{\underline{U}_t \geq 0}{\text{minimum}} & F_t(\underline{U}_t) \\ \text{subject to} & M_t\underline{U}_t = \delta_t\underline{X} \\ & N_t\underline{U}_t \geq Q_t^* \end{array} \right| \right\} \tag{7.7}$$

where the optimization within the inner brackets is just the subproblem discussed above. The set Ω consists of all vectors $\underline{X}$ which allow a feasible solution in each of the subproblems. Since the problem is convex, as will be demonstrated in the next section, the inner optimization can be replaced by its nonlinear programming dual

$$\underset{\substack{\underline{\lambda}_t \\ \pi_t \geq 0}}{\text{maximize}} \ \underset{\underline{U}_t \geq 0}{\text{minimum}} \ [F_t(\underline{U}_t) + \underline{\lambda}_t(M_t\underline{U}_t - \delta_t\underline{X}) - \pi_t(N_t\underline{U}_t - Q_t^*)] \tag{7.8}$$

where $\underline{\lambda}_t$ is a vector of dual multipliers, or shadow prices, on the loading order constraints (7.2), and π_t is a scalar multiplier on the peak-load constraint (7.3).

The feasible set Ω can be equivalently described as the set of vectors $\underline{X}$ for which the maximum (7.8) is finite. Since an infinite value can be obtained when

$$N_t\underline{U}_t < Q_t^*$$

$$\text{or} \quad M_t\underline{U}_t \neq \delta_t\underline{X}$$

this condition is equivalent to

$$\underset{\underline{U}_t \geq 0}{\text{minimum}} \; [\underline{\mu}_t(M_t\underline{U}_t - {}_t\underline{X}) - \pi_t(N_t\underline{U}_t - Q_t^*)] \leq 0$$

$$\text{for all} \quad \pi_t \geq 0 \quad \text{and} \quad \underline{\mu}_t$$

Then, the generalized Benders' master problem can be written

$$\underset{\substack{Z \\ \underline{X} \geq 0}}{\text{maximize}} \; Z$$

$$\text{subject to} \quad Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} \min_{\underline{U}_t \geq 0} \; [F_t(\underline{U}_t) + \underline{\lambda}_t(M_t\underline{U}_t - \delta_t\underline{X})$$

$$-\pi_t(N_t\underline{U}_t - Q_t^*)] \quad \text{for all} \quad \pi_t \geq 0 \text{ and } \underline{\lambda}_t \qquad (7.9)$$

$$0 \geq \min_{\underline{U}_t \geq 0} \left[\underline{\mu}_t(M_t \underline{U}_t - \delta_t \underline{X}) - \pi_t(N_t \underline{U}_t - Q_t^*)\right]$$

$$t = 1, \ldots, T, \text{ for all } \pi_t \geq 0 \text{ and } \underline{\mu}_t \qquad (7.10)$$

$\underline{X} \geq 0$, Z unrestricted in sign

Solving this problem is equivalent to solving the original problem (7.1) - (7.3); however, it has an infinite number of constraints since the number of possible values for $\underline{\lambda}_t$, $\underline{\mu}_t$ and π_t is infinite. It is generally solved by the strategy of relaxation, in which the constraints (7.9) and (7.10) are generated successively. A relaxed version of the problem which includes only a few of the constraints (7.9) and (7.10) is solved to find a trial solution $\hat{Z}$, $\hat{\underline{X}}$. If this trial solution violates some of the constraints not yet included, then one or more of the violated constraints is generated and joined to the relaxed problem, which is solved again to find a new trial solution. This procedure is continued until a trial solution is generated which satisfies all of the ignored constraints, and thus is optimal, or until a solution which is acceptably close to optimality has been found.

Given a trial solution $\hat{\underline{X}}$, a violated constraint of the form (7.9) can be generated by solving the problems, for $t = 1, \ldots, T$,

$$\max_{\substack{\pi_t \geq 0 \\ \underline{\lambda}_t}} \min_{\underline{U}_t \geq 0} [F_t(\underline{U}_t) + \underline{\lambda}_t(M_t\underline{U}_t - \delta_t\hat{\underline{X}}) - \pi_t(N_t\underline{U}_t - Q_t^*)] \qquad (7.11)$$

which are the nonlinear programming duals of the operating subproblems for periods $t = 1,\ldots,T$

$$\underset{\underline{U}_t \geq 0}{\text{minimize}} \; F_t(\underline{U}_t)$$

$$\text{subject to} \quad M_t\underline{U}_t = \delta_t\hat{\underline{X}}$$

$$N_t\underline{U}_t \geq Q_t^*$$

The solution obtained from solving equations (7.5) is $\hat{\underline{U}}_t$. The optimal dual multipliers $\hat{\pi}_t$ and $\hat{\underline{\lambda}}_t$ are determined from the Kuhn-Tucker conditions for this problem. The details of this solution are discussed in Section B of this chapter. The cost constraint thus generated in the master problem is

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [F_t(\underline{U}_t) + \hat{\underline{\lambda}}_t(M_t\hat{\underline{U}}_t - \delta_t\underline{X}) - \pi_t(N_t\hat{\underline{U}}_t - Q_t^*)] \quad \text{or}$$

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [F_t(\hat{\underline{U}}_t) + \hat{\underline{\lambda}}_t\delta_t(\hat{\underline{X}} - \underline{X})]$$

since $M_t\hat{\underline{U}}_t = \delta_t\hat{\underline{X}}$ and since complementary slackness guarantees that $\hat{\pi}_t(N_t\hat{\underline{U}}_t - Q_t^*) = 0$. This constraint is the <u>most</u> <u>violated</u> among those of (7.9) which have not yet been included in the relaxed master problem because it was chosen by maximizing the right-hand sides of these constraints. If the trial solution $\hat{Z}$, $\hat{\underline{X}}$ satisfies this constraint then it must also satisfy all of the other as-yet-unincluded constraints.

If the trial solution $\hat{\underline{X}}$ leads to an infeasible subproblem in some periods t, a violated constraint of the form (7.10) can be generated by solving the subproblems

$$\max_{\substack{\underline{\mu}_t \\ \pi_t \geq 0}} \ \min_{\underline{U}_t \geq 0} \ [\underline{\mu}_t(M_t\underline{U}_t - \delta_t\hat{\underline{X}}) - \pi_t(N_t\underline{U}_t - Q_t^*)].$$

If, for any values of $\pi_t \geq 0$ and $\underline{\mu}_t$, the inner minimization yields a positive value, then the maximum can be increased without bound simply by multiplying $\underline{\mu}_t$ and π_t by $\alpha > 0$ and letting $\alpha \to +\infty$. For definiteness, choose $\pi_t = 1$. Then this problem is the linear programming dual of

$$\underset{\underline{U}_t \geq 0}{\text{minimize}} \ N_t\underline{U}_t$$

$$\text{subject to } M_t\underline{U}_t = \delta_t\hat{\underline{X}} \tag{7.12}$$

and the optimal multipliers $\hat{\underline{\mu}}_t$ are determined from the dual problem. Further discussion of this problem is found in Section C of this chapter. The feasibility constraint thus generated in the master problem is

$$\underline{\mu}_t(M_t\underline{\hat{U}}_t-\delta_t\underline{X}) - (N_t\underline{\hat{U}}_t-Q_t^*) \geq 0$$

$$\text{or} \quad N_t\underline{\hat{U}}_t+\underline{\hat{\mu}}_t\delta_t(X-\hat{X}) \geq Q_t^*$$

It will be shown in Section C of this chapter that the optimal dual solution for the problem is $\underline{\hat{\mu}}_t = \underline{e} \equiv [1,1,\ldots,1]'$. Since

$$N_t\underline{\hat{U}}_t = \hat{U}^{I_t t} = \sum_{i=1}^{I_t} \hat{X}^{it} = \underline{e}'\delta_t\underline{\hat{X}},$$

this constraint can be written

$$\underline{e}'\delta_t\underline{X} \geq Q_t^*$$

$$\text{or} \quad \sum_{i=1}^{I_t} X^{it} \geq Q_t^* \;;$$

that is, enough capacity must be available in each period t to satisfy the peak demand. Clearly, these constraints are so simple that they can be included in the master problem from the beginning, rather than being generated when infeasible subproblems are discovered.

Finally, the relaxed masterproblem can be written

$$\begin{aligned}
&\underset{Z,\ \underline{X}}{\text{minimize}} \quad Z \\
&\text{subject to} \quad Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [F_t(\underline{U}_t^k) + \underline{\lambda}_t^k \delta_t (\underline{X}^k - \underline{X})] \qquad k = 1, \ldots, K \\
&\qquad\qquad \underline{e}'\delta_t \underline{X} \geq Q_t^* \qquad t = 1, \ldots, T \\
&\qquad\qquad\quad \underline{X} \geq 0
\end{aligned}$$

where $\underline{X}^k$ is the k^{th} trial solution generated and $\underline{U}_t^k$ and $\underline{\lambda}_t^k$ are the associated primal and dual solutions to the operating subproblems. This relaxed master problem is just the problem (3.13) - (3.15) given in Chapter 3. Notice that it is a linear program.

Since the procedure described above generates the most violated of the master problem constraints, in the sense that if the current trial solution $\hat{\underline{X}}$, $\hat{Z}$ satisfies the new constraint it generates, it satisfies all the remaining constraints, the current trial solution is optimal when it satisfies the newly generated constraint. Also, since the current trial solution $\hat{\underline{X}}$ is a feasible capacity expansion plan, the cost of this plan (the right-hand side of the new constraint, with $\underline{X} = \hat{\underline{X}}$) is an upper bound on the optimal cost. Furthermore, the current value $\hat{Z}$ is a lower bound, since it is determined in a relaxed version of the master problem. These bounds can be used to terminate the procedure prior to optimality with known error bounds. If a satisfactory solution (optimal or near-optimal) has not been found, then the new constraints are added to the master problem, and a new trial plan is determined as the optimal solution to the new relaxed master.

B. Solution of the Subproblem

In the previous section, the operating subproblem (7.4) - (7.6) and its dual (7.8) were used to generate cost constraints in the master problem. In order to replace the subproblem by its dual, the global optimality conditions, described in the introduction to Part Two, must be satisfied.

Proposition: The operating subproblem (7.4) - (7.6) is a convex program.

Proof: The load duration function $G(Q)$ decreases monotonically with increasing load since the time when the load exceeds a higher level Q_1 is a subset of the time in which it exceeds a lower level Q_2. Thus

$$G(Q_1) \leq G(Q_2) \quad \text{for} \quad Q_1 \geq Q_2.$$

If Q^* is the peak load during the period, then $G(Q) = 0$ for $Q \geq Q^*$. Furthermore, $G(Q)$ is defined to be constant (equal to the duration of the period) for $Q \leq 0$.

Define the function

$$W(Q) = \int_0^Q G(\xi)d\xi$$

in each subproblem. Since $G(Q)$ decreases monotonically,

$W(Q)$ is concave. Since $G(Q) = 0$ for $Q \geq Q^*$, $W(Q) = W(Q^*)$ for $Q \geq Q^*$. The objective function of the subproblem (7.4) can be rewritten

$$\sum_{i=1}^{I} F^i[W(U^i) - W(U^{i-1})] =$$

$$F^I W(U^I) - F^1 W(0) - \sum_{i=1}^{I-1} [F^{i+1} - F^i] W(U^i)$$

By definition of the merit order, $F^{i+1} \geq F^i$. Furthermore, $W(0) = 0$ and because of the peak-load constraint (7.6), $W(U^I) = W(Q^*)$. Hence the objective function is equivalent to

$$F^I W(Q^*) - \sum_{i=1}^{I-1} [F^{i+1} - F^i] W(U^i) \qquad (7.13)$$

where the first term is constant and the remaining terms are convex. Therefore, the functions $F_t(\underline{U}_t)$ are convex, and the operating subproblem for each period is a convex program, since its feasible region is defined by linear inequalities (7.5) and (7.6). ||

The convexity of this problem means that a solution which satisfies the Kuhn-Tucker conditions, Theorem 1 of the Introduction, is globally optimal, and the Strong Duality property, Theorem 3, allows replacement of the subproblem by its dual.

The operating subproblem can be rewritten

$$\text{minimize} \quad F^I W(Q^*) - \sum_{i=1}^{I-1} [F^{i+1} - F^i] W(U^i) \qquad (7.13)$$

$$\text{subject to} \quad U^i - U^{i-1} = X^i \qquad i = 1,\ldots,I \qquad (7.5)$$

$$U^I \geq Q^* \qquad (7.6)$$

$$U^i \geq 0$$

The problem (7.8) is just the nonlinear programming dual of this subproblem. Let λ^i be the dual multiplier associated with the $i^{\underline{th}}$ loading order constraint (7.5) and π be the multiplier associated with the peak-load constraint (7.6). Since constraints (7.5) are equalities, λ^i is unrestricted in sign while π must be non-negative. Define the Lagrangian function for the problem $L(\underline{U},\underline{\lambda},\pi)$ in the usual manner. The Kuhn-Tucker conditions for optimality in this problem are

i) $\dfrac{\partial L}{\partial U^i} \geq 0$ with equality if $U^i > 0$;

ii) Complementary slackness between π and constraint (7.6); and

iii) $U^1,\ldots,U^I$ satisfy constraints (7.5) and (7.6).

Note that because the problem is convex, as was proved above, these conditions are sufficient, as well as necessary, for optimality.

Assuming, for the moment, that this subproblem is feasible and that $U^i > 0$ for all i (which is guaranteed by having $X^1 > 0$), condition (i) gives a set of equations

$$\lambda^i - \lambda^{i+1} = (F^{i+1} - F^i)\, G(U^i) \qquad i = 1,\ldots,I-1 \qquad (7.14)$$

$$\lambda^I - \pi = 0$$

$$\pi \geq 0,\ \lambda^i \text{ unrestricted in sign}$$

If, as will often be the case, $U^I > Q^*$, then $\pi = 0$, and the shadow prices λ^i can be determined by solving this set of backward recursion equations. If $U^I = Q^*$, which will almost certainly occur when the set of plant capacities is optimal in the planning problem, or if $U^i = 0$ for some of the plants i, which can occur if some of the plants low in the merit order are uneconomical to build, then the subproblems are degenerate, and the Kuhn-Tucker conditions give only inequalities instead of equations. These degenerate cases are discussed in Section D of this chapter.

As has been noted previously, the constraints (7.5) have a single solution

$$U^i = \sum_{n=1}^{i} X^n \qquad i = 1,\ldots,I \qquad (7.15)$$

and this solution is feasible if $U^I \geq Q^*$. Thus solving the operating subproblem is almost trivially easy. The plants are arranged in merit order, and equations (7.15) are used to compute the loading points U^i. Then equations (7.14) are used to find the shadow prices which generate the constraint in the master problem. Even though the subproblem is a nonlinear program, no explicit optimization is required to solve it, because the optimal solution is known beforehand to be merit order operation. The explicit optimization is performed in the master problem, which is a linear program.

C. Subproblem Infeasibility

It was assumed in the previous section that each subproblem has a feasible solution. Insuring feasibility is a matter of generating the proper constraints on the values of X^i which are sent to the subproblem. These constraints are implicit in the subproblem, and when a violation occurs, it is necessary to generate an explicit constraint on the $\underline{X}$ in the master problem.

Consider the related subproblem

$$\underset{U^1,\ldots,U^I}{\text{maximize}} \quad U^I$$

$$\text{subject to} \quad U^i - U^{i-1} = X^i \qquad i = 1,\ldots,I$$

$$U^i \geq 0$$

which is just problem (7.12) of Section A. Clearly, the operating subproblem (7.4)-(7.6) has a feasible soltuion if and only if there is a solution to this related subproblem with

$$Q^* \leq \max U^I$$

This problem is a linear program which always has a feasible solution, namely

$$U^I = X^1 + \dots + X^I.$$

The dual of this problem is

$$\underset{\mu^1,\dots,\mu^I}{\text{minimize}} \quad \sum_{i=1}^{I} X^i \mu^i$$

$$\text{subject to} \quad \mu^i - \mu^{i+1} \geq 0 \qquad i = 1,\dots,I-1$$

$$\mu^I \geq 1$$

$$\mu^i \text{ unrestricted in sign}$$

which can be restated as

$$\text{minimize} \quad \sum_{i=1}^{I} X^i \mu^i$$

$$\text{subject to} \quad \mu^1 \geq \mu^2 \geq \dots \geq \mu^I \geq 1$$

If all $X^i > 0$, then clearly the unique optimal solution is to set

$$\mu^1 = \mu^2 = \dots = \mu^I = 1$$

and then

$$\max U^I = X^1 + \dots + X^I$$

by the duality theorem for linear programs. The solution is unique except when

$$X^1 = X^2 = \ldots = X^I = 0.$$

Thus the master problem constraint which insures feasibility in the operating subproblem for period t is

$$\sum_{i=1}^{I_t} X^{it} \geq Q_t^*,$$

which is the feasibility constraint used in the master problem in Section A. While this constraint is intuitively obvious, the rigorous derivation serves as a demonstration of the technique, which will be used in deriving feasibility constraints for the probabilistic model in the next chapter, where they are not obvious.

D. Subproblem Degeneracy

Consider now what happens if in the subproblem (7.4)-(7.6)

$$U^I = Q^*.$$

In this case, it is not necessary that $\pi = 0$ and therefore, the Kuhn-Tucker conditions do not give a unique set of multipliers. Therefore, this case represents a problem of degeneracy. Suppose that λ_0^i is the solution to the recursion equations (7.14) obtained when $\pi = 0$. In the degenerate case, any other solution of the form

$$\lambda^i = \lambda_0^i + \pi \qquad i = 1,\ldots,I$$

with $\pi \geq 0$ is also valid.

In this degenerate case, the dual subproblem becomes

$$\max_{\substack{\lambda \\ \pi \geq 0}} \min_{\underline{U} \geq 0} \{F(\underline{U}) + \sum_{i=1}^{I} (\lambda_0^i + \pi)(U^i - U^{i-1} - X^i) - \pi(U^I - Q^*)\}$$

$$= \max_{\substack{\lambda \\ \pi \geq 0}} \min_{\underline{U} \geq 0} \{F(\underline{U}) + \sum_{i=1}^{I} \lambda_0^i (U^i - U^{i-1} - X^i) + \pi(Q^* - \sum_{i=1}^{I} X^i\}$$

which is finite if and only if $\sum_{i=1}^{I} X^i \geq Q^*$. This is just the feasibility constraint derived in the previous section.

The meaning of this argument can be elucidated by the following reasoning. In the degenerate case, if the value of Q^* is perturbed slightly to $Q^*+\delta$, the degeneracy disappears. If δ is negative, then the situation is just the nondegenerate case discussed in Section C, and the optimal multipliers λ^i are unique, as given by equation (7.14). If δ is positive, however, the subproblem becomes infeasible, since now

$$U^I = X^1 + \ldots + X^I < Q + \delta.$$

In this case, a feasibility constraint would be generated, according to the reasoning of Section C. Hence, a degenerate subproblem must generate both a set of shadow prices λ_0^i and a feasibility constraint.

A similar argument can be applied in the degenerate case when $U^i = 0$ for some set of i. Recall that in this case, the Kuhn-Tucker conditions are inequalities in the shadow prices, rather than equations, so that again the λ^i are not unique. Applying the perturbation argument to the plant capacities X^i generates a set of shadow prices λ_0^i by considering the Kuhn-Tucker conditions as equalities, when $\delta > 0$, and a feasibility constraint when $\delta < 0$. The feasibility constraint is just

$$X^i \geq 0.$$

CHAPTER 8

SOME TECHNICAL ISSUES ASSOCIATED WITH THE PROBABILISTIC MODEL

A. Derivation of the Master Problem

Recall that the probabilistic capacity expansion planning model is stated in Chapter 4 as follows:

$$\underset{\underline{X},\underline{Y},\ldots,\underline{Y}_T}{\text{minimize}} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t) \tag{8.1}$$

$$\text{subject to} \quad EG_t(\underline{Y}_t) \leq \varepsilon_t \tag{8.2}$$

$$\underline{Y}_t \leq \delta_t X \qquad t = 1,\ldots,T \tag{8.3}$$

$$\underline{X} \geq 0 \quad \underline{Y}_t \geq 0$$

where, as before, $\underline{X}$ is the vector of plant capacities to be built, and $\underline{C}$ is the vector of plant capital costs per unit of capacity. For period t, the matrix δ_t creates the merit ordering of the plant capacities, and the vector $\underline{Y}_t$ contains the utilization levels of the plants. The constraining value for unserved energy in period t is ε_t.

Associated with this capacity planning problem, there is, for each period t in the planning horizon, an operating subproblem of the form

$$\underset{Y^1,\ldots,Y^I}{\text{minimize}} \quad \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)\,dQ \tag{8.4}$$

$$\text{subject to} \quad \int_{U^I}^{\infty} G_{I+1}(Q)\,dQ \le \varepsilon \tag{8.5}$$

$$0 \le Y^i \le X^i \tag{8.6}$$

(The index t indicating the time period has been suppressed for clarity.) As before, Y^i is the utilization level for the i^{th} plant, and the plant loading points are defined by

$$U^i - U^{i-1} = Y^i \qquad i = 1,\ldots,I \tag{8.7}$$

with $U^0 = 0$. The function $G_i(Q)$ represents the equivalent load duration curve faced by the i^{th} plant, defined by the probabilistic simulation recursion

$$G_{i+1}(Q) = p_i G_i(Q) + q_i G_i(Q-Y^i) \qquad i = 1,\ldots,I \tag{8.8}$$

with $G_1(Q) \equiv G(Q)$, where $G(Q)$ is the system load duration curve, p_i is the availability of plant i, and its forced outage rate is $q_i = 1-p_i$. The operating cost coefficients, as before, are denoted by F^i, and the merit order indices i are defined so that $F^i \le F^{i+1}$.

These operating subproblems are included within the capacity planning problem. The plant utilization levels Y^i

in the subproblem for period t form the vector $\underline{Y}_t$ in the capacity planning problem. The objective function (8.4) is the operating cost function $EF_t(\underline{Y}_t)$, and the expected unserved energy in (8.5) is represented by the function $EG_t(\underline{Y}_t)$.

The probabilistic capacity planning model (8.1)-(8.3) can be written in equivalent form as a two-stage optimization

$$\underset{\substack{\underline{X} \geq 0 \\ \underline{X} \in \Omega}}{\text{minimize}} \left\{ \underline{C}'\underline{X} + \sum_{t=1}^{T} \left| \begin{array}{l} \underset{\underline{Y}_t \geq 0}{\text{minimum}} \; EF_t(\underline{Y}_t) \\ \\ \text{subject to } EG_t(\underline{Y}_t) \leq \varepsilon_t \\ \qquad\qquad\qquad \underline{Y}_t \leq \delta_t \underline{X} \end{array} \right| \right\} \tag{8.9}$$

where the optimization within the inner brackets is just the operating subproblem (8.4)-(8.6). The set Ω consists of all capacity vectors $\underline{X}$ which allow a feasible solution in each of the subproblems. By the duality properties of the subproblem, to be discussed in the next section, the inner optimization can be replaced by its dual.

$$\begin{aligned} \underset{\underline{\lambda}_t, \pi_t \geq 0}{\text{maximize}} \; \underset{\underline{Y}_t \geq 0}{\text{minimum}} \; \{ & EF_t(\underline{Y}_t) + \pi_t [EG_t(\underline{Y}_t) - \varepsilon_t] \\ & + \underline{\lambda}_t [\underline{Y}_t - \delta_t \underline{X}] \} \end{aligned} \tag{8.10}$$

where $\underline{\lambda}_t$ is a vector of dual multipliers on the capacity constraints (8.6) and π_t is a scalar multiplier on the unserved energy constraint (8.5).

The feasible set Ω can be equivalently described as the set of vectors $\underline{X}$ for which the maximum in (8.10) is finite. Since an infinite value can be obtained for the maximization if any component of $[\underline{Y}_t - \delta_t \underline{X}]$ is positive or if $EG_t(\underline{Y}_t) > \varepsilon_t$, this condition is equivalent to

$$\underset{\underline{Y}_t \geq 0}{\text{minimum}} \{\pi_t [EG_t(\underline{Y}_t) - \varepsilon_t] - \underline{\mu}_t [\underline{Y}_t - \delta_t \underline{X}]\} \leq 0$$

$$\text{for all } \pi_t, \underline{\mu}_t \geq 0$$

Then, the generalized Benders' master problem can be written

$$\underset{Z,\underline{X}}{\text{minimize}} \; Z$$

subject to

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} \min_{\underline{Y}_t \geq 0} \{EF_t(\underline{Y}_t) + \pi_t [EG_t(\underline{Y}_t) - \varepsilon_t] + \underline{\lambda}_t [\underline{Y}_t - \delta_t \underline{X}]\} \quad (8.11)$$

$$\text{for all } \pi_t, \underline{\lambda}_t \geq 0$$

$$0 \geq \sum_{t=1}^{T} \min_{\underline{Y}_t \geq 0} \{\pi_t[EG_t(U_t)-\varepsilon_t] + \underline{\mu}_t[\underline{Y}_t-\delta_t\underline{X}]\} \quad (8.12)$$

for all $\pi_t, \underline{\mu}_t \geq 0$

$\underline{X} \geq 0$

Solving this problem is equivalent to solving the original problem (8.1) - (8.3).

This master problem is solved by the relaxation strategy, described in the previous chapter, in which the constraints (8.11) and (8.12) are generated successively. A relaxed version of the problem, containing only a few of these constraints, is solved for a trial solution $(\hat{Z},\hat{\underline{X}})$. A violated constraint of the form (8.11) can be generated by solving the problem, for each period t,

$$\underset{\underline{\lambda}_t,\pi_t \geq 0}{\text{maximize}} \ \underset{\underline{Y}_t \geq 0}{\text{minimize}} \{EF_t(\underline{Y}_t) + \pi_t[EG_t(\underline{Y}_t)-\varepsilon_t] + \underline{\lambda}_t[\underline{Y}_t-\delta_t\hat{\underline{X}}]\}.$$

As noted above, this problem is the dual of the operating subproblem

$$\underset{\underline{Y}_t \geq 0}{\text{minimize}} \ EF_t(\underline{Y}_t)$$

$$\text{subject to} \ EG_t(\underline{Y}_t) \leq \varepsilon_t$$

$$\underline{Y}_t \leq \delta_t\hat{\underline{X}}$$

The optimal primal solution $\hat{\underline{Y}}_t$ is obtained by solving this subproblem, and the optimal dual multipliers $\hat{\pi}_t$ and $\hat{\underline{\lambda}}_t$ are determined from the Kuhn-Tucker conditions for this problem. The details of solving the subproblems are discussed in Section B of this chpater. The <u>cost</u> constraint thus generated in the master problem is

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} \{EF_t(\hat{\underline{Y}}_t) + \hat{\pi}_t[EG_t(\hat{\underline{Y}}_t)-\varepsilon_t] + \hat{\underline{\lambda}}_t[\hat{\underline{Y}}_t-\delta_t\underline{X}]\}$$

or $Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [EF_t(\hat{\underline{Y}}_t) + \hat{\underline{\lambda}}_t\delta_t(\hat{\underline{X}}-\underline{X})]$ since

$\hat{\pi}_t[EG_t(\hat{\underline{Y}}_t)-\varepsilon_t] = 0$ and $\hat{\underline{\lambda}}_t(\hat{\underline{Y}}_t-\delta_t\hat{\underline{X}}) = 0$ by complementary slackness.

If the trial solution $\hat{\underline{X}}$ leads to infeasible subproblems in some of the periods t, a violated constraint of the form (8.12) is generated by solving the subproblems

$$\underset{\pi_t,\underline{\mu}_t}{\text{maximize}}\ \underset{\underline{Y}_t\geq 0}{\text{minimum}}\ \{\pi_t[EG_t(\underline{Y}_t)-\varepsilon_t] + \underline{\mu}_t[\underline{Y}_t-\delta_t\hat{\underline{X}}]\}.$$

If, for any value of π_t, $\underline{\mu}_t \geq 0$, the inner minimization yields a positive value, then the maximum can be increased without bound. By normalizing, as in the previous chapter, with $\pi_t = 1$, this problem becomes the dual of

$$\underset{\underline{Y}_t \geq 0}{\text{minimize}} \quad EG_t(\underline{Y}_t)$$

$$\text{subject to} \quad \underline{Y}_t \leq \delta_t \hat{\underline{X}}$$

The optimal primal solution $\hat{\underline{Y}}_t$ is obtained by solving this problem and the optimal multipliers $\underline{\mu}_t$ are determined from the Kuhn-Tucker conditions. The details of the solution are discussed in Section C of this chapter. The <u>feasibility</u> constraint generated in the master problem is

$$0 \geq \sum_{t \varepsilon \Gamma} \{[EG_t(\hat{\underline{Y}}_t) - \varepsilon_t] + \hat{\underline{\mu}}_t[\hat{\underline{Y}}_t - \delta_t \underline{X}]\}$$

or, since by complementary slackness $\hat{\underline{\mu}}_t[\hat{\underline{Y}}_t - \delta_t \hat{\underline{X}}] = 0$,

$$\sum_{t \varepsilon \Gamma} [EG_t(\hat{\underline{Y}}_t) + \hat{\underline{\mu}}_t \delta_t (\hat{\underline{X}} - \underline{X})] \leq \sum_{t \varepsilon \Gamma} \varepsilon_t$$

where the sums are taken over the set Γ of periods in which $EG_t(\hat{\underline{Y}}_t) > \varepsilon_t$.

Finally, the relaxed master problem can be written

$$\underset{Z,\underline{X}}{\text{minimize}} \quad Z$$

subject to

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [EF_t(\underline{Y}_t^k) + \underline{\lambda}_t^k \delta_t(\underline{X}^k - \underline{X})] \quad k = 1,\ldots,K$$

$$\sum_{t\epsilon\Gamma^k} [EG_t(\underline{Y}_t^k) + \underline{\mu}_t^k \delta_t(\underline{X}^k - \underline{X})] \leq \sum_{t\epsilon\Gamma^k} \epsilon_t \quad k = 1,\ldots,K$$

$$\underline{X} \geq 0$$

where $\underline{X}^k$ is the $k^{\underline{th}}$ trial solution generated and $\underline{Y}_t^k$ and $\underline{\lambda}_t^k$, $\underline{\mu}_t^k$ are the associated primal and dual solutions to the operating subproblems. This relaxed master problem is just the problem given in Chapter 4, (4.16) - (4.18). Notice that, as in the last chapter, this master problem is a linear program.

As noted previously, the constraints generated by solving the subproblems are the most violated of the constraints not yet included in the master problem. Hence if the current trial solution $(\hat{Z},\hat{\underline{X}})$ satisfies the new constraints it generates, then it is optimal. Furthermore, the value of $\hat{Z}$ is a lower bound on the cost of the optimal solution, and if $\hat{\underline{X}}$ is feasible, the cost of this solution is an upper bound on the optimal cost. These bounds can be used to terminate the procedure prior to optimality with known error bounds. If a satisfactory solution (optimal or near-optimal) has not been found, then the newly generated constraints are added to the master problem, which is then re-solved to generate a new trial plan.

B. Solution of the Subproblem

The subproblem of the probabilistic planning model is somewhat more complex than that of the deterministic model discussed in the previous chapter. The optimal solution of this subproblem is not quite so obvious as it was in the deterministic case, nor are its duality properties established by a simple convexity argument. Nevertheless, it will be shown that this subproblem is a relatively well-behaved optimization problem and that the required duality properties do hold. The strategy of this section will be to propose solutions for the subproblem and its dual and to show that they satisfy the global optimality conditions stated in the introduction to this part. Not only will this argument demonstrate that the proposed solutions are, in fact, optimal, but it will also establish the Strong Duality property required to justify replacing the subproblem by its dual in the derivation of the generalized Benders' master problem.

The operating subproblem of the probabilistic model was stated above as

$$\underset{Y^1,\ldots,Y^I}{\text{minimize}} \quad EF(\underline{Y}) = \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)\,dQ \tag{8.4}$$

$$\text{subject to} \quad EG(\underline{Y}) = \int_{U^I}^{\infty} G_{I+1}(Q)\,dQ \le \varepsilon \tag{8.5}$$

$$0 \le Y^i \le X^i \tag{8.6}$$

$$\text{where} \quad U^i - U^{i-1} = Y^i \quad \text{with} \quad U^0 = 0 \tag{8.7}$$

Assume, for the moment, that this subproblem has a feasible solution and that $X^i > 0$ for all i. This problem is not a convex program, since the cost function $EF(\underline{Y})$ is not convex.

The proposed optimal solution for this problem is to set $Y^i = X^i$ successively for each plant in merit order until the unserved energy constraint (8.5) is exactly satisfied. The last plant so loaded, the <u>marginal</u> plant, will generally not have to be used at full capacity. The plants above the marginal plant do not operate. Let n be the index of the marginal plant; then the solution is

$$\hat{Y}^i = \begin{cases} X^i & \text{for} \quad i < n \\ 0 & \text{for} \quad i > n \end{cases} \tag{8.13}$$

and $\hat{Y}^n$ is set so that

$$\int_{U^n}^{\infty} G_{n+1}(Q)dQ = \varepsilon$$

with $\quad 0 < \hat{Y}^n \leq X^n$

The Kuhn-Tucker conditions can be applied to this solution to propose a solution to the dual problem. Let π be the dual multiplier associated with the unserved energy constraint (8.5), and let λ^i be the multiplier associated with the $i^{\underline{th}}$ capacity constraint (8.6). Since all of the constraints are inequalities, π and λ^i must all be non-negative. Define the Lagrangian function for the problem as

$$L(\underline{Y},\underline{\lambda},\pi) = EF(\underline{Y}) + \pi[EG(\underline{Y})-\varepsilon] + \underline{\lambda}[\underline{Y}-\underline{X}]$$

$$= \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ + \pi[\int_{U^I}^{\infty} G_{I+1}(Q)dQ-\varepsilon]$$

$$+ \sum_{i=1}^{I} \lambda^i[Y^i-X^i] \tag{8.14}$$

The Lagrangian minimization problem associated with the subproblem is to

$$\underset{\underline{Y} \geq 0}{\text{minimize}}\ L(\underline{Y},\underline{\lambda},\pi)$$

with the multipliers π and $\underline{\lambda}$ held fixed as parameters. The Kuhn-Tucker conditions give necessary conditions for a <u>local</u> minimum in this problem:

i) Stationarity: $\frac{\partial}{\partial Y^i} L(\underline{Y},\underline{\lambda},\pi) \geq 0 \quad i = 1,\ldots,I$

with equality if $Y^i > 0$

ii) Complementary Slackness:

$$\pi[EG(\underline{Y})-\varepsilon] = 0$$

$$\underline{\lambda}[\underline{Y}-\underline{X}] = 0$$

Then, dual multipliers satisfying these conditions when $\underline{Y} = \hat{\underline{Y}}$ are

$$\hat{\lambda}^i = \begin{cases} -\frac{\partial}{\partial Y^i} EF(\hat{\underline{Y}}) - \hat{\pi}\frac{\partial}{\partial Y^i} EG(\hat{\underline{Y}}) & \text{for } i < n \\ 0 & \text{for } i \geq n \end{cases}$$

Now, $\hat{\pi}$ is the shadow price on unserved energy, the marginal cost associated with a small decrease in ε. Since decreasing the unserved energy requires additional generation from the marginal plant, this marginal cost is just the cost of operating that plant. Hence

$$\hat{\pi} = F^n \tag{8.16}$$

This reasoning can be confirmed rigorously by referring to the Kuhn-Tucker conditions for plant n. Assuming $0 < \hat{Y}^n < X^n$ (the degenerate case when equality holds at one of these bounds is discussed in Section D), then $\hat{\lambda}^n = 0$ and condition (i) must hold with equality. Therefore

$$\hat{\pi}\frac{\partial}{\partial Y^n}EG(\hat{\underline{Y}}) = -\frac{\partial}{\partial Y^n}EF(\hat{\underline{Y}})$$

or, since $\hat{Y}^i = 0$ for $i > n$,

$$\hat{\pi}\frac{\partial}{\partial Y^n}\int_{U^n}^{\infty} G_{n+1}(Q)dQ = -F^n p_n \frac{\partial}{\partial Y^n}\int_{U^{n-1}}^{U^n} G_n(Q)dQ.$$

Evaluating the derivatives gives

$$\hat{\pi}(-p_n G_n(U^n)) = -F^n p_n G_n(U^n)$$

or $\hat{\pi} = F^n$ as promised. A similar argument shows that for $i > n$

$$\hat{\pi}\frac{\partial}{\partial Y^i}EG(\hat{\underline{Y}}) \geq -\frac{\partial}{\partial Y^i}EF(\hat{\underline{Y}})$$

as required by condition (i).

These proposed solutions $\hat{\underline{Y}}$ and $(\hat{\underline{\lambda}},\hat{\pi})$ can be proven optimal by showing that they satisfy the global optimality conditons stated in the Introduction to this part:

i) Minimality: $\hat{\underline{Y}}$ minimizes $L(\underline{Y},\hat{\underline{\lambda}},\hat{\pi})$ over $\underline{Y} \geq 0$.

ii) Complementary Slackness.

iii) Feasibility.

The feasibility condition (iii) is satisfied because $\hat{\underline{Y}}$ is feasible in (8.4) - (8.6) and because $\underline{\lambda} \geq 0$ and $\hat{\pi} \geq 0$. The complementary slackness condition (ii) is the same as appears in the Kuhn-Tucker conditions and is therefore satisfied (recall that $\hat{Y}^n$ is defined so that there is no slack in the unserved energy constraint). Hence, it is only required to show that the minimality condition (i) is satisfied. Now, the Kuhn-Tucker stationarity condition is a necessary condition for a local minimum, but it is not sufficient. The proof that $\hat{\underline{Y}}$ minimizes $L(\underline{Y},\hat{\underline{\lambda}},\hat{\pi})$ actually consists of two propositions. The first establishes that at least part of the Lagrangian is convex; the second establishes that at a local minimum of the Lagrangian, the nonconvex part must be zero. Then, by the well-known property of convex functions that a stationary point is a global minimum, it can be concluded that the minimality condition (i) is satisfied.

<u>First Proposition</u>: The function

$$\sum_{i=1}^{n} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ + \hat{\pi} \int_{U^n}^{\infty} G_{n+1}(Q)dQ \qquad (8.17)$$

is convex.

Proof: Despite the apparent complexity of the function, resulting from the recursive definition of the equivalent load duration functions $G_i(Q)$, it actually has a rather simple structure. It is a convex combination of simpler functions defined in various outage states.

Define an outage state σ to be the occurrence of a random event in which a given subset I_σ of the plants are available and the complementary subset I_σ' are out. Given I plants in all, there are 2^I possible outage states. Define the indicators

$$\theta_{i\sigma} = \begin{cases} 1 & \text{if plant i is available in state } \sigma\text{, and} \\ 0 & \text{if not.} \end{cases}$$

The probability of state σ occurring is

$$\phi_\sigma = \prod_{i=1}^{I} p_i^{\theta_{i\sigma}} q_i^{(1-\theta_{i\sigma})} .$$

Define the plant loading points in state σ as

$$V_\sigma^i - V_\sigma^{i-1} = \theta_{i\sigma} Y^i \qquad i = 1,\dots,I$$

with $V_\sigma^0 = 0$. As in the deterministic case, the energy produced by plant i in state σ is

$$\int_{V_\sigma^{i-1}}^{V_\sigma^i} G(Q)\,dQ$$

which is zero if plant i is not available because $V_\sigma^i = V_\sigma^{i-1}$. Notice that the system load duration curve is used rather than the equivalent load duration curve. For plants which are available in state σ, an alternative expression for the energy produced is

$$\int_{U^{i-1}}^{U^i} G(Q - U^{i-1} + V_\sigma^{i-1})\,dQ$$

where the term $U^{i-1} - V_\sigma^{i-1}$ is the capacity on outage below plant i in the merit order.

The equivalent load duration curve G_i is an expectation over all outage states

$$G_i(Q) = \sum_{\sigma=1}^{2^I} \phi_\sigma G(Q - U^{i-1} + V_\sigma^{i-1}).$$

Because the number of outage states grows exponentially with the number of plants, it is not convenient to deal with the states explicitly. Defining the equivalent load duration curves recursively, as was done in Chapter 4, allows the outage states to be dealt with implicitly.

Similarly, the expected energy produced by plant i is

$$\sum_{\sigma=1}^{2^I} \phi_\sigma \int_{V_\sigma^{i-1}}^{V_\sigma^i} G(Q)\,dQ = \sum_\sigma \phi_\sigma \int_{U^{i-1}}^{U^i} G(Q-U^{i-1}+V_\sigma^{i-1})\,dQ$$

where the second summation is made over all states σ in which plant i is available. This last sum is equivalent to

$$p_i \int_{U^{i-1}}^{U^i} G_i(Q)\,dQ,$$

the expression for expected energy which was derived previously.

Now, if $\hat{\pi}$ is considered to be a cost coefficient for unserved energy, then the cost for operating plants up to and including plant n in state σ is

$$\sum_{i=1}^{n} F^i \int_{V_\sigma^{i-1}}^{V_\sigma^i} G(Q)\,dQ + \hat{\pi} \int_{V_\sigma^n}^{\infty} G(Q)\,dQ \tag{8.18}$$

This function is convex in the V_σ^i, and therefore in the Y^i, by the same argument used to establish the convexity of the cost function in the deterministic model (in Section B of Chapter 7). That is, the function

$$W(Q) = \int_0^Q G(\xi)\,d\xi$$

is concave because $G(Q)$ is monotonically decreasing. Thus (8.18) can be rewritten as

$$\sum_{i=1}^{n} F^i[W(V_\sigma^i)-W(V_\sigma^{i-1})]+\hat{\pi}[W(\infty)-W(V_\sigma^n)]$$

$$= \hat{\pi}W(\infty) - \sum_{i=1}^{n-1}[F^{i+1}-F^i]W(V_\sigma^i) - [\hat{\pi}-F^n]W(V_\sigma^n)$$

which is convex because $F^{i+1} \geq F^i$ by definition of the merit order, $\hat{\pi} = F^n$, and $W(Q)$ is concave.

Now, the function (8.17) is just the expectation, taken over all outage states,

$$\sum_{\sigma=1}^{2^I} \phi_\sigma[\sum_{i=1}^{n} F^i \int_{V_\sigma^{i-1}}^{V_\sigma^i} G(Q)dQ + \hat{\pi}\int_{V_\sigma^n}^{\infty} G(Q)dQ]$$

$$= \sum_{i=1}^{n} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ + \hat{\pi}\int_{U^n}^{\infty} G_{n+1}(Q)dQ,$$

which is a convex combination of convex functions because

$$\sum_{\sigma=1}^{2^I} \phi_\sigma = 1 \text{ and } \phi_\sigma \geq 0$$

and is therefore convex. ||

<u>Second</u> <u>Proposition</u>: In the Lagrangian minimization problem, with $\underline{\lambda} = \hat{\underline{\lambda}}$ and $\pi = \hat{\pi}$, any local minimum has

$Y^i = 0$ for $i > n$.

Proof: Suppose there is a solution with $Y^i > 0$ for some $i > n$. In fact, let i index the plant highest in the merit order with $Y^i > 0$. Then

$$U^I = U^i$$

and $$G_{I+1}(Q) = G_{i+1}(Q),$$

and the terms of the Lagrangian (8.14) which depend on Y^i are

$$\hat{\lambda}^i Y^i + F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ + \hat{\pi} \int_{U^i}^{\infty} G_{i+1}(Q)dQ$$

This expression can be rewritten as

$$\hat{\lambda}^i Y^i + F^i p_i \int_{U^{i-1}}^{U^{i-1}+Y^i} G_i(Q)dQ + \hat{\pi} p_i \int_{U^{i-1}+Y^i}^{\infty} G_i(Q)dQ$$

$$+ \pi q_i \int_{U^{i-1}}^{\infty} G_i(Q)dQ$$

using the recursive definition of G_{i+1} in (8.8); the final term does not depend on Y^i. Since $G_i(Q)$ is positive, the function

$$W_i(Q) = \int_0^Q G_i(Q)dQ$$

is increasing. Then the Y^i dependent terms in the expression above can be written as

$$\hat{\lambda}^i Y^i + p_i[F^i - \pi]W_i(U^{i-1}+Y^i) + \pi p_i W_i(\infty) - F^i p_i W_i(U^{i-1})$$

where the final two terms do not depend on Y^i. Since $\hat{\pi} = F^n \leq F^i$ and $\hat{\lambda}^i = 0$, the terms

$$p_i[F^i - \hat{\pi}]W_i(U^{i-1}+Y^i)$$

decrease as Y^i decreases. Hence, the cost of the solution with $Y^i > 0$ can be reduced by taking $Y^i = 0$. This argument can be applied successively to each $Y^i > 0$, $i > n$, to show that $Y^i = 0$ for all $i > n$ in any locally optimal solution to the Lagrangian minimization problem. ||

These two propositions together establish that $\hat{Y}$ minimizes $L(\underline{Y}, \hat{\underline{\lambda}}, \hat{\pi})$ over $\underline{Y} \geq 0$. For the second proposition establishes that in any local minimum $Y^i = 0$ for $i > n$. Therefore minimizing the Lagrangian (8.14) is equivalent to minimizing the function

$$\sum_{i=1}^{n} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ + \hat{\pi} \int_{U^n}^{\infty} G_{n+1}(Q)dQ + \hat{\lambda}^i Y^i.$$

This function is the sum of (8.17), proved convex in the first proposition, and a linear term. It is therefore convex, and hence, the stationarity condition (i) of the Kuhn-Tucker conditions is sufficient to establish global minimality. Thus, $\hat{\underline{Y}}$ satisfies the global optimality condition (i), and $\hat{\underline{Y}}$ and $(\hat{\underline{\lambda}},\hat{\pi})$ together constitute the optimal primal and dual solutions for the subproblem (8.4) - (8.6).

An important consequence of this proof is that the optimal value of the objective function in the primal subproblem (8.4) - (8.6) is equal to the optimal value of the dual problem

$$\underset{\underline{\lambda},\pi}{\text{maximize}}\ \underset{\underline{Y}\,\underline{\geq}\,0}{\text{minimum}}\ L(\underline{Y},\underline{\lambda},\pi),$$

the Strong Duality property. This fact justifies the use of the dual problem in place of the primal in the derivation of the master problem in Section A.

Actual computation of the solution to the subproblem and, particularly, of the subproblem shadow prices is a sufficiently intricate calculation to warrant a separate discussion, found in Chapter 9. However, while determination of these solutions is not quite as trivial as it was in the deterministic model, discussed in the previous chapter, it still does not require any explicit optimization. Hence the separation of the explicit optimization in the master problem, which is a linear program, from the implicit optimization in the subproblem leads to an efficient algorithm for solving the probabilistic capacity planning problem.

C. Subproblem Infeasibility

It was assumed in the previous section that each subproblem has a feasible solution. As noted before, insuring feasibility is a matter of generating the proper constraints on the values of $\underline{X}$ which are sent to the subproblems. These constraints are implicit in the subproblems, and when a violation occurs, it is necessary to generate an explicit constraint on the $\underline{X}$ in the master problem.

Consider the problem of determining, for given $\underline{X}$, whether or not subproblems have feasible solutions. This problem can be posed as an optimization of the form

$$\underset{\substack{W_1,\ldots,W_T \\ \underline{Y}_1,\ldots,\underline{Y}_T}}{\text{minimize}} \quad \sum_{t=1}^{T} W_t \tag{8.19}$$

$$\text{subject to} \quad EG_t(\underline{Y}_t) - \varepsilon_t = W_t \tag{8.20}$$

$$\underline{0} \leq \underline{Y}_t \leq \delta_t \underline{X} \qquad t = 1,\ldots,T \tag{8.21}$$

$$W_t \geq 0$$

The variables W_t are called <u>artificial</u> variables, and the objective function is usually known as an <u>infeasibility form</u>. Clearly the vector $\underline{X}$ allows feasible solutions if and only if the minimum in this problem is zero.

The problem (8.19) - (8.21) is a convex program, and it is equivalent to the dual problem

$$\underset{\substack{\pi_1,\dots,\pi_T \\ \underline{\mu}_1,\dots,\underline{\mu}_T\geq 0}}{\text{maximize}} \quad \underset{\substack{W_1,\dots,W_T\geq 0 \\ \underline{Y}_1,\dots,\underline{Y}_T\geq 0}}{\text{minimum}} \quad \sum_{t=1}^{T} \{W_t + \pi_t[EG_t(\underline{Y}_t)-\varepsilon_t-W_t]$$

$$+ \underline{\mu}_t[\underline{Y}_t-\delta_t\underline{X}]\} \qquad (8.22)$$

This dual problem must have its maximum equal to zero to insure feasibility. Note that the dual problem decomposes into independent problems for each period, of the form

$$\underset{\substack{\pi_t \\ \underline{\mu}_t\geq 0}}{\text{maximize}} \quad \underset{\substack{W_t\geq 0 \\ \underline{Y}_t\geq 0}}{\text{minimum}} \quad \{W_t+\pi_t[EG_t(\underline{Y}_t)-\varepsilon_t-W_t]+\mu_t[\underline{Y}_t-\delta_t\underline{X}]\}$$

The optimal dual multipliers π and $\underline{\mu}$ for this problem can be found by applying the Kuhn-Tucker conditions, which require that

i) $1 - \pi \geq 0$, with equalify if $W > 0$

ii) $\pi\frac{\partial}{\partial Y^i}EG(\underline{Y}) + \mu^i \geq 0$, with equality if $Y^i > 0$.

iii) $\pi[EG(\underline{Y})-\varepsilon-W] = 0$

$\underline{\mu}[\underline{Y}-\underline{X}] = 0$

If the subproblem is infeasible for this period, then $W > 0$ and $\pi = 1$. Thus the dual problem given above is equivalent to

$$\text{minimize} \quad EG_t(\underline{Y}_t)$$
$$\text{subject to} \quad 0 \leq \underline{Y}_t \leq \delta_t \underline{X}$$

which was given in Section A. In order to minimize the infeasibility, each Y^i will be set to its upper bound X^i. Hence (assuming that $X^i > 0$), the multipliers μ^i are determined by

$$\mu^i = -\frac{\partial}{\partial Y^i} EG(\underline{Y}) \qquad i = 1,\ldots,I \tag{8.23}$$

$$\text{with} \quad Y^i = X^i \qquad i = 1,\ldots,I$$

If, on the other hand, the problem is feasible, then an optimal solution is to set $\pi = 0$ and $\mu_i = 0$.

Now, for a particular trial solution $\hat{\underline{X}}$, solving the problem (8.19) - (8.21) generates particular values $\hat{W}_t$, $\hat{\underline{Y}}_t$, $\hat{\underline{\mu}}_t$, and $\hat{\pi}_t$ which satisfy the dual relationship above. Then for any feasible value of $\underline{X}$ it must be true that

$$\sum_{t=1}^{T} \{\hat{W}_t+\hat{\pi}_t[EG_t(\hat{\underline{Y}}_t)-\varepsilon_t-\hat{W}_t] + \hat{\underline{\mu}}_t[\hat{\underline{Y}}_t-\delta_t\underline{X}]\} \leq 0$$

or $$\sum_{t\in\Gamma} \{EG_t(\hat{\underline{Y}}_t)-\varepsilon_t+\hat{\underline{\mu}}_t\delta_t[\hat{\underline{X}}-\underline{X}]\} \leq 0$$

where Γ is the set of periods in which $\hat{\underline{X}}$ gives infeasible subproblems. This is the feasibility constraint given in Section A.

It should be noted that even when the subproblem is infeasible, it is legitimate to generate a cost constraint as well. In this case, all the plants will be used to capacity, so that $Y^i = X^i$ for all i, and thus the marginal plant will be the last plant in the merit order, so that $n = I$ and $\pi = F^I$.

D. Subproblem Degeneracy

When the subproblem has degenerate solutions, the Kuhn-Tucker conditions do not uniquely specify the dual multipliers. Degeneracy can occur in two situations - first, if $X^i = 0$ so that Y^i satisfies both its upper and lower bounds simultaneously, and second, if $Y^n = X^n$, so that the marginal plant is used to capacity. In both cases, the degeneracy is resolved by a perturbation argument.

The first case, where $X^i = 0$, is relatively straightforward. Suppose X^i were to be perturbed to a small positive value. If $i < n$, that is plant i is below the marginal plant n in the merit order, it would be optimal to increase Y^i to a positive value as well. On the other hand, if $i > n$, it would be optimal to leave Y^i at a zero level. Hence, in this case, Y^i behaves as if it were at its capacity when $i < n$ and as if it were at zero when $i > n$. Therefore, the dual multipliers are calculated in exactly the same way as if X^i were positive instead of zero.

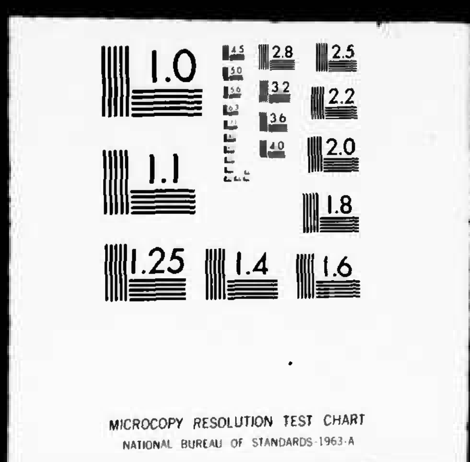
The second case, where $Y^n = X^n$, is less straightforward. Suppose X^n were to be perturbed by a small amount to $X^n + \delta$. If $\delta > 0$, then nevertheless, it would not be optimal to increase Y^n. Therefore, plant n would remain the marginal plant and $\pi = F^n$. If, on the other hand, $\delta < 0$, then Y^n would have to decrease and Y^{n+1} would increase from zero. Then plant n+1 would become the marginal plant and $\pi = F^{n+1}$. Hence, in this case, any value of π between F^n and F^{n+1} would be acceptable as optimal, and alternative values of λ^i could be calculated with different values of π. The dual solution is not unique.

AD-A058 421 MASSACHUSETTS INST OF TECH CAMBRIDGE OPERATIONS RESE--ETC F/G 12/2
DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY --ETC(U)
AUG 78 J A BLOOM DAAG29-76-C-0064
UNCLASSIFIED TR-154 ARO-14261.10-M NL

3 OF 4
AD
A058421

3 OF 4

AD A058421



CHAPTER 9
COMPUTATIONAL METHODS FOR THE PROBABILISTIC OPERATING SUBPROBLEM

A. Introduction

Although the optimal solution to the operating subproblem of the probabilistic model and its associated shadow prices were derived in the previous chapter, the actual computation of the solution was not discussed. In this chapter, some computational methods will be presented.

The optimal solution to the subproblem was shown, in the last chapter, to be

$$Y^i = \begin{cases} X^j & \text{for} \quad 1 \leq j < n \\ 0 & \text{for} \quad n < j \leq I \end{cases} \tag{9.1}$$

where the marginal plant index n is defined so that

$$\int_{U^n}^{\infty} G_{n+1}(Q)\,dQ = \varepsilon$$

with $0 < Y^n \leq X^n$. The optimal dual multipliers, or shadow prices, were shown to be

$$\lambda^j = \begin{cases} -\frac{\partial}{\partial Y^j}EF(\underline{Y}) - \pi\frac{\partial}{\partial Y^j}EG(\underline{Y}) & \text{for } 1 \le j < n \\ 0 & \text{for } n \le j \le I \end{cases} \tag{9.2}$$

and $\pi = F^n$.

If the subproblem is infeasible, then $Y^j = X^j$ for all j, $\pi = F^I$, and the multipliers for the feasibility constraint were shown to be

$$\mu^j = -\frac{\partial}{\partial Y^j}EG(\underline{Y}) \qquad j = 1,\ldots,I \tag{9.3}$$

Recall that the expected operating cost function is defined by

$$EF(\underline{Y}) = \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q)dQ \tag{9.4}$$

and the expected unserved energy function by

$$EG(\underline{Y}) = \int_{U^I}^{\infty} G_{I+1}(Q)dQ \tag{9.5}$$

The equivalent load duration curves are defined by the probabilistic simulation recursion

$$G_{i+1}(Q;Y^1,\ldots,Y^i) = p_i G_i(Q;Y^1,\ldots,Y^{i-1})$$
$$+ q_i G_i(Q-Y^i;Y^1,\ldots,Y^{i-1}) \qquad (9.6)$$

where $p_i + q_i = 1$. The arguments after the semi-colon have been included to show explicitly the dependence on Y. (Call the argument before the semi-colon the zeroth argument, those after, the first through $i^{\underline{th}}$.) The plant loading points defined by

$$U^i - U^{i-1} = Y^i \qquad i = 1,\ldots,I$$
$$\text{with } U^0 = 0 \qquad (9.7)$$

All of the required computations in solving the subproblem are based on the probabilistic simulation recursion (9.6). The procedure is to convolve, one by one, each plant of the merit order into the load duration curve to form the equivalent load duration curve for the next plant. Using the equivalent load duration curve for the $i^{\underline{th}}$ plant, it is determined whether this plant is the marginal plant, and its contribution to the expected operating cost and to meeting the unserved energy constraint are calculated, as will be described in Section B. The contribution of the i^{th} load duration curve to the multipliers λ^j and

μ^j are also computed, as will be described in Section C. Determining the marginal plant is the key to finding the utilization levels of the plants, since according to (9.1), each plant below the marginal plant is used to capacity while plants above it are not used. Once the marginal plant has been loaded, the probabilistic simulation recursion can be terminated, since the remaining plants, loaded at zero level, do not change the equivalent load duration curve.

It is worthwhile, before concluding this section, to establish some properties of the equivalent load duration functions $G_i(Q)$. Consider first some properties of the system load duration function $G(Q)$:

i) $G(Q) = 1$ for $Q < 0$

ii) $G(Q) = 0$ for $Q \geq Q^*$

iii) $G(Q)$ is monotonically decreasing

These properties are inherited by the equivalent load duration functions, as can easily be shown by induction using the recursive definition (9.6):

i) $G_i(Q) = 1$ for $Q < 0$

ii) $G_i(Q) = 0$ for $Q \geq Q^* + Y^1 + \ldots + Y^{i-1}$

iii) $G_i(Q)$ is monotonically decreasing

In a computer program implementing these calculations, the function $G_i(Q)$ is actually represented by a vector giving the function's values at discrete points of its domain. The number of points required to represent the function, and hence the dimension of the vector, depends on the size of the increments into which the domain is divided. Values of the function at points between those represented in the vector can be found by interpolation, and values of its integral can be computed numerically. It is not, however, necessary to represent the cost function EF or the unserved energy function EG in this way since their values will be computed only at the optimal utilization levels $\underline{Y}$.

B. Determining the Marginal Plant and Computing Expected Operating Cost and Energy

Finding the marginal plant is the controlling factor in the probabilistic simulation loop, since once it has been found, the loop can be terminated. The marginal plant is defined as that plant which just satisfies the unserved energy constraint. Therefore, as each plant is loaded in the probabilistic simulation loop, its contribution to meeting the unserved energy is accumulated, until the plant which just satisfies the constraint is found.

Let $\Delta E_i(Y^i)$ be the expected energy produced by the i^{th} plant when it is loaded at level Y^i

$$\Delta E_i(Y^i) = \int_{U^{i-1}}^{U^{i-1}+Y^i} G_i(Q)\,dQ \qquad (9.8)$$

Note that the unserved energy constraint

$$\int_{U^I}^{\infty} G_{I+1}(Q)\,dQ \leq \varepsilon$$

can be rewritten

$$\sum_{i=1}^{I} \int_{U^{i-1}}^{U^i} G_i(Q)\,dQ \geq E - \varepsilon$$

where E is the total energy demanded

$$E = \int_0^{Q^*} G(Q)\,dQ$$

because the total energy demand must equal the sum of the expected energy produced and the expected unserved energy. Now suppose $n-1$ plants have been loaded and the unserved energy constraint has not yet been satisfied. Since all of these plants will have been loaded to capacity, the n^{th} plant will satisfy the constraint if and only if

$$\Delta E_n(Y^n) + \sum_{i=1}^{n-1} \Delta E_i(X^i) = E - \varepsilon \qquad (9.9)$$

for some Y^n such that $0 < Y^n \leq X^n$. When the energy $\Delta E_n(Y^n)$ is computed by numerically integrating (9.8), using, for example, the trapezoidal rule, it is computed for successive increments of Y^n between zero and X^n. Thus, it is a simple matter to detect a value for Y^n for which (9.9) is satisfied. If such a value is found, then plant n is, of course, the marginal plant. If not, then Y^n is set equal to X^n and the search for the marginal plant must continue. In either case, the contribution of this plant to the expected operating cost is

$$F^n p_n \Delta E_n(Y^n)$$

and to meeting the unserved energy constraint is

$$\Delta E_n(Y^n).$$

Furthermore, once the utilization level of the plant has been determined, it can be convolved using (9.6) to determine the equivalent load duration curve for the next plant.

C. Computation of the Shadow Prices

A large part of the work required to solve the probabilistic planning model is involved with computing the dual multipliers (9.2) and (9.3) in the subproblems. The complexity of this computation is a result of the recursive definition of the equivalent load duration curves $G_i(Q)$. Since a great deal of computation must be invested in computing these functions anyway, it might be hoped that they alone would be sufficient to compute the desired multipliers. Unfortunately, this does not seem to be the case. There does not seem to be a simple, non-recursive relationship between the multipliers and the equivalent load duration curves. The computational method presented in this section, however, appears to be reasonably efficient.

Calculating the dual multipliers in (9.2) and (9.3) requires computation of the derivatives

$$\frac{\partial}{\partial Y^j} \int_{U^{i-1}}^{U^i} G_i(Q;Y^1,\ldots,Y^{i-1})\,dQ$$

The computation is trivial for $i \leq j$ since G_i does not depend on Y^j for $i \leq j$ and U^i does not depend on Y^j for $i < j$. Hence

$$\frac{\partial}{\partial Y^j}\int_{U^{i-1}}^{U^i} G_i(Q;Y^1,\ldots,Y^{i-1})dQ = \begin{cases} 0 & \text{for } i < j \\ G_j(U^i) & \text{for } i = j \end{cases}$$

Now let U stand for any U^i with $i > j$ so that $\frac{\partial}{\partial Y^j}U = 1$. Define the function

$$H_{ij}(U) = \frac{\partial}{\partial Y^j}\int_0^U G_i(Q;Y^1,\ldots,Y^{i-1})dQ$$

for all $i > j$. When $i = j + 1$, then

$$H_{j+1,j}(U) = \frac{\partial}{\partial Y^j}\int_0^U G_{j+1}(Q;Y^1,\ldots,Y^j)dQ$$

$$= \frac{\partial}{\partial Y^j}\int_0^U p_jG_j(Q;Y^1,\ldots,Y^{j-1})dQ + \frac{\partial}{\partial Y^j}\int_0^U q_jG_j(Q-Y^j;Y^1,\ldots,Y^{j-1})dQ$$

$$= \frac{\partial}{\partial Y^j}\int_0^U p_jG_j(Q;Y^1,\ldots,Y^{j-1})dQ + \frac{\partial}{\partial Y^j}\int_{-Y^j}^{U-Y^j} q_jQ_j(Q;Y^1,\ldots,Y^{j-1})dQ$$

$$= p_jG_j(U) + \text{a constant}$$

because $U - Y^j$ does not depend on Y^j. The constant of integration does not depend on U and will cancel out when differences are taken below. Thus

$$\frac{\partial}{\partial Y^j}\int_{U^j}^{U^{j+1}} G_{j+1}(Q;Y^1,\ldots,Y^j)dQ = H_{j+1,j}(U^{j+1}) - H_{j+1,j}(U^j)$$

In general, when $i \geq j + 1$

$$H_{i+1,j}(U) = \frac{\partial}{\partial Y^j} \int_0^U G_{i+1}(Q;Y^1,\ldots,Y^i)dQ$$

$$= \frac{\partial}{\partial Y^j} \int_0^U p_i G_i(Q;Y^1,\ldots,Y^{i-1})dQ + \frac{\partial}{\partial Y^j} \int_0^U q_i G_i(Q-Y^i;Y^1,\ldots,Y^{i-1})dQ$$

$$= p_i H_{ij}(U) + q_i H_{ij}(U-Y^i) + \text{a constant}$$

$$\text{and} \quad \frac{\partial}{\partial Y^i} \int_{U^{i-1}}^{U^i} G_i(Q;Y^1,\ldots,Y^{i-1})dQ = H_{ij}(U^i) - H_{ij}(U^{i-1})$$

where the constants of integration cancel when the differences are taken. In summary, the recursive definition for the H_{ij} functions is the same as that of the G_i functions with a different initial condition

$$\begin{aligned} H_{i+1,j}(U) &= p_i H_{ij}(U) + q_i H_{ij}(U-Y^i) \quad i = j+1,\ldots,I \\ H_{j+1,j}(U) &= p_j G_j(U) \end{aligned} \tag{9.10}$$

The initial condition, which defines $H_{j+1,j}$, arises from the "transfer" of Y^j from the j^{th} argument to the zeroth argument in the definition of G_{j+1}. Notice in this calculation, no actual integration or differentiation of the G_i functions need be performed.

Now, the derivatives of the expected cost and unserved energy functions can be written

$$\frac{\partial}{\partial Y^j} EF(\underline{Y}) = F^j p_j G_j(U^j) + \sum_{i=j+1}^{I} F^i p_i [H_{ij}(U^i) - H_{ij}(U^{i-1})]$$

$$\frac{\partial}{\partial Y^j} EG(\underline{Y}) = - H_{I+1,j}(U^I) \tag{9.11}$$

The following properties of the H_{ij} functions are easily demonstrated by induction using their recursive definition (9.10) and the properties of the G_i functions:

i) $H_{ij}(U) = p_j$ for $U < 0$

ii) $H_{ij}(U) = 0$ for $U \geq Q^* + Y^1 + \ldots + Y^{i-1} - Y^j$

iii) $H_{ij}(U)$ is monotonically decreasing.

The recursive definition given in (9.10) would seen to be an inefficient way to compute the H_{ij} functions, for two reasons. First, the definition is actually doubly recursive, once on i and once on j; hence there is a great deal more work involved in computing the H_{ij} than in computing the G_i alone. Second, the H_{ij} functions must be computed over their entire domain, even though only their values at one or two points are actually used in computing the required derivatives; the values at the other points are required only at later stages of the recursion. In a computer program implementing this calculation, each function H_{ij} would be represented by a vector giving its values at discrete points of its domain. Although $H_{i+1,j}$ could be written over H_{ij} in a memory table, there would still have to be a separate vector in the table for each j, using a great deal of memory.

There is an alternative to the recursive definition (9.10) to compute the H_{ij}. Since, as was shown above

$$H_{j+1,j}(U) = p_j G_j(U),$$

applying the recursion for G_{j+1} gives the identity

$$p_j H_{j+1,j}(U) + q_j H_{j+1,j}(U-Y^j) = p_j G_{j+1}(U).$$

Furthermore, since the recursions for H_{ij} and G_i have the same form, applying this recursion successively to both sides of this identity gives

$$p_j H_{ij}(U) + q_j H_{ij}(U-Y^j) = p_j G_i(U) \quad i = j+1,\dots,I+1 \qquad (9.12)$$

Now, assuming $Y^j > 0$ and $p_j > 0$, solving (9.12) for $H_{ij}(U)$ gives

$$H_{ij}(U) = G_i(U) - \frac{q_j}{p_j} H_{ij}(U-Y^j) \tag{9.13}$$

Furthermore, repeatedly applying (9.13)

$$H_{ij}(U-Y^j) = G_i(U-Y^j) - \frac{q_j}{p_j} H_{ij}(U-2 \cdot Y^j)$$

$$\vdots$$

$$H_{ij}(U-k \cdot Y^j) = G_i(U-k \cdot Y^i) - \frac{q_j}{p_j} H_{ij}(U-(k+1) \cdot Y^j)$$

Now, if k is defined so that

$$k \cdot Y^j < U \leq (k+1) \cdot Y^j$$

then $\quad H_{ij}(U-(k+1) \cdot Y^j) = p_j$

by property (i) of the function H_{ij} given above. Hence, if $Y^j > 0$, then

$$H_{ij}(U) = \sum_{\ell=0}^{k} \left(-\frac{q_j}{p_j}\right)^{\ell} G_i(U-\ell \cdot Y^j) + \left(-\frac{q_j}{p_j}\right)^{k+1} p_j \tag{9.14}$$

and if $Y^j = 0$, (9.12) reduces to

$$H_{ij}(U) = p_j G_i(U) \tag{9.15}$$

If Y^j is larger than the interval between points in the discrete representation of $H_{ij}(U)$ (as it most likely would be), the computation of (9.14) requires fewer steps than the computation of (9.10), since (9.14) computes points of H_{ij} at intervals of Y^j rather than at the interval of the discrete representation as (9.10) does. Furthermore, computation of (9.14) requires much less memory than computation of (9.10) since it computes $H_{ij}(U)$ directly from $G_i(U)$, so that there is no need to store the H_{ij} functions.

The calculation of the derivatives (9.11) can be further simplified by eliminating the need to calculate $H_{ij}(U^{i-1})$. For, repeatedly applying (9.10) gives

$$H_{ij}(U^{i-1}) = p_{i-1}H_{i-1,j}(U^{i-1}) + q_{i-1}p_{i-2}H_{i-2,j}(U^{i-2}) + \ldots$$
$$+ q_{i-1}q_{i-2}\ldots q_{j+2}p_{j+1}H_{j+1,j}(U^{j+1}) + q_{i-1}q_{i-2}\ldots q_{j+1}p_jG_j(U^j).$$

Define the probability factors in this expression as

$$\beta_{i-1,k} = q_{i-1}\ldots q_{k+1}$$

or recursively

$$\begin{aligned} \beta_{ii} &= 1 \\ \beta_{ik} &= \beta_{i,k+1}q_{k+1} \end{aligned} \tag{9.16}$$

Then $H_{ij}(U^{i-1}) = \beta_{i-1,j}p_jG_j(U^j) + \sum_{k=j+1}^{i-1} \beta_{i-1,k}p_kH_{kj}(U^k)$.

Thus

$$\frac{\partial}{\partial Y^j}EF = F^ip_jG_j(U^i) + \sum_{i=j+1}^{I} F^ip_j[H_{ij}(U^i) - H_{ij}(U^{i-1})]$$

$$= [F^j - \sum_{i=j+1}^{I} \beta_{i-1,j}p_iF^i]p_jG_j(U^j)$$

$$+ \sum_{i=j+1}^{I} F^ip_i[H_{ij}(U^i) - \sum_{k=j+1}^{i-1} \beta_{i-1,k}p_kH_{kj}(U^k)]$$

$$= [F^j - \sum_{i=j+1}^{I} \beta_{i-1,j}p_iF^i]p_jG_j(U^j)$$

$$+ \sum_{k=j+1}^{I} [F^k - \sum_{i=k+1}^{I} \beta_{i-1,k}p_iF^i]p_kH_{kj}(U^k)$$

Define the coefficient

$$\Delta_j = F^j - \sum_{i=j+1}^{I} \beta_{i-1,j}p_iF^i$$

This coefficient satisfies the backward recursion

$$\Delta_{j-1} = F^{j-1} - F^j + q_j\Delta_j \quad j = I,\ldots,2 \tag{9.17}$$

where $\Delta_I = F^I$.

Expanding the definition gives

$$\Delta_j = F^j - p_{j+1}F^{j+1} - q_{j+1}p_{j+2}F^{j+2} - \dots - q_{j+1}q_{j+2}\dots q_{I-1}p_I F^I$$

from which the significance of this coefficient can be seen. It is the expected difference in operating cost between plant j and the next available plant in the merit order.

Thus finally, the cost function derivative can be written

$$\frac{\partial}{\partial Y^j}EF = \Delta_j p_j G_j(U^j) + \sum_{i=j+1}^{I} \Delta_i p_i H_{ij}(U^i) \qquad (9.18)$$

which requires computation of H_{ij} at only the single point U^i. The expected unserved energy derivative can be further rearranged in similar fashion, although there would seem to be no computational advantage.

$$\frac{\partial}{\partial Y^j}EG = - H_{I+1,j}(U^I)$$

$$= \beta_{Ij}p_j G_j(U^i) - \sum_{i=j+1}^{I} \beta_{Ii}p_i H_{ij}(U^i) \qquad (9.19)$$

One final simplification can be made, using the fact that $Y^i = 0$ for $i > n$ in the optimal solution to the subproblem. Hence

$$G_i(Q) = G_{n+1}(Q) \quad \text{for} \quad i \geq n + 1$$

and $$U^i = U^n \quad \text{for} \quad i \geq n$$

From (9.10), it is also true that

$$H_{ij}(U) = H_{n+1,j}(U) \quad \text{for} \quad i \geq n+1,\ j \leq n$$

Hence

$$\frac{\partial}{\partial Y^j}EF = \Delta_j p_j G_j(U^j) + \sum_{i=j+1}^{n} \Delta_i p_i H_{ij}(U^i)$$

$$+ \left(\sum_{i=n+1}^{I} \Delta_i p_i\right) H_{n+1,j}(U^n) \quad \text{for} \quad j \leq n$$

By the definition (9.17) of Δ_i

$$p_i \Delta_i = F^i - F^{i+1} + q_{i+1}\Delta_{i+1} - q_i \Delta_i$$

and $p_I \Delta_I = F^I - q_I \Delta_I$.

Hence

$$\sum_{i=n+1}^{I} \Delta_i p_i = F^I + \sum_{i=n+1}^{I-1} (F^i - F^{i+1})$$

$$- q_I \Delta_I + \sum_{i=n+1}^{I-1} (\Delta_{i+1} q_{i+1} - \Delta_i q_i)$$

$$= F^{n+1} - \Delta_{n+1} q_{n+1}.$$

Therefore

$$\frac{\partial}{\partial Y^j}EF = \Delta_j p_j G_j(U^i) + \sum_{i=j+1}^{n} \Delta_i p_i H_{ij}(U^i)$$

$$+ (F^{n+1} - \Delta_{n+1} q_{n+1}) H_{n+1,j}(U^n)$$

Now finally, the computation of the shadow prices (9.2) can be written

$$\pi = F^n$$

$$\lambda^j = -\frac{\partial}{\partial Y^j} EF - \pi \frac{\partial}{\partial Y^j} \quad \text{for} \quad j < n$$

$$= -\Delta_j p_j G_j(U^j) - \sum_{i=j+1}^{n} \Delta_i p_i H_{ij}(U^i)$$

$$- (F^{n+1} - \Delta_{n+1} q_{n+1}) H_{n+1,j}(U^n) + F^n H_{n+1,j}(U^j)$$

$$= -\Delta_j p_j G_j(U^j) - \sum_{i=j+1}^{n} \Delta_i p_i H_{ij}(U^i) + \Delta_n H_{n+1,j}(U^n) \qquad (9.20)$$

$$\text{for} \quad j < n$$

$$\lambda^j = 0 \quad \text{for} \quad j \geq n$$

$$\mu^j = -\frac{\partial}{\partial Y^j} EG = H_{I+1,j}(U^I)$$

(Since μ^j are required when the subproblem is infeasible, $n = I$ in computing these coefficients.)

D. Outline of the Complete Algorithm

Based on the discussions in the previous sections of this chapter, a complete algorithm can be proposed for finding the optimal solution to the subproblem and its associated dual multipliers. This algorithm is outlined below.

Initially: The loading point U is set to zero, the equivalent load duration curve is set equal to the system load duration curve, and the unserved energy is set equal to the total energy demanded, the area under the system load duration curve. The coefficients for expected difference in cost, Δ_i, are calculated for all plants using the backward recursion (9.17).

For each plant i in the merit order: Entering this loop, the current value of the equivalent load duration curve is G_i and the current loading point U is U^{i-1}. If the unserved energy constraint has not yet been satisfied, then the expected energy generated by plant i is computed by numerically integrating the current load duration curve from the current loading point U to the utilization level Y^i. The integration is performed by incrementing Y^i in small steps. If the unserved energy constraint is satisfied at some value of Y^i less than or equal to the capacity of the plant X^i, then i is the marginal plant n, and the utilization level is left at this value; otherwise, Y^i is set equal to the capacity X^i. The loading point is increased to $U + Y^i$ and the unserved energy is decreased by the amount of energy generated by the plant.

The terms of the shadow prices which depend on the current load point and equivalent load duration curve are computed and added to the accumulating values for these variables. These terms are

$$\text{for } i < n,\ \lambda^i \leftarrow -\Delta_i p_i G_i(U^i)$$

$$\text{for } i \leq n \text{ and all } j < i,\ \lambda^j \leftarrow \lambda^i - \Delta_i p_i H_{ij}(U^i)$$

$$\text{for } i = n+1 \text{ and all } j < n,\ \lambda^j \leftarrow \lambda^j - \Delta_n H_{n+1,j}(U^n)$$

$$\text{for } i = I+1 \text{ and all } j \leq I,\ \mu^j \leftarrow H_{I+1,j}(U^I)$$

For each $j < i$, the term $H_{ij}(U)$ is computed using the recursion (9.14) or (9.15).

If the marginal plant has not yet been loaded, the next equivalent load duration curve is computed using the recursion (9.6), and it replaces the current load duration curve. The plant index i is incremented and the algorithm repeats the computations within this loop, until either the marginal plant has been found and loaded or until all the plants in the merit order have been loaded.

Finally: If all the plants have been loaded and the unserved energy constraint remains unsatisfied, then the subproblem is infeasible and the algorithm generates a feasibility constraint for the master problem based on the multipliers μ^j. Whether the problem is feasible or infeasible, a cost constraint for the master problem is generated, based on the multipliers λ^j.

Part Three

Application of the Decomposition Approach to Peak-Load Pricing

Introduction

The purpose of Part Three is to show how the decomposition approach to electric utility capacity planning models, introduced in the previous chapters of this thesis, can be extended to the problem of determining peak-load prices for electricity. Peak-load pricing, or time-of-day pricing, as it is sometimes called, is based on the economic principle of marginal cost. Since the cost of generating electricity depends on the load, which in turn varies by time of day, users should be charged different prices for electricity consumed at different times. Intuitively, this variation in cost arises from two factors. First, in merit order operation, plants with higher operating costs are brought into service as the load increases. Second, increases in the peak load may require the purchase of additional generating capacity, thus incurring additional capacity costs. Because the highest costs are incurred in periods of peak demand, peak-load pricing generally requires a higher price in these periods than in off-peak periods when the demand is low.

Peak-load pricing has received a great deal of attention in the economic research literature in the past, and recently interest has become more widespread as regulatory bodies in many states move toward implementation of peak-load tariffs. Implementation of peak-load pricing has been encouraged by a number of factors. Consumer groups have been pressing for more equitable electric rates. Since the "Energy Crisis" has brought scarcity and higher prices for many fuels, particularly petroleum, ways have been sought to encourage conservation and more efficient use of these fuels. Because of financial difficulties for the industry in the past few years, electric utilities have been trying to increase utilization of their existing plants in order to reduce their need for new construction. Peak-load pricing has been proposed as a remedy for these problems.

Traditional work on peak-load pricing has developed in two distinct directions (see the recent survey by Joskow[1]). One approach has emphasized the demand side, particularly price elasticity and shifts in the time pattern of consumption. This approach has often used simple supply models, with only a single generating technology or a single plant. The other approach has emphasized the supply side, particularly estimation of marginal costs. This approach has often used more complex supply models, but has de-emphasized demand-related issues. More recently, there has been work toward integrating the two approaches.

The approach taken in this thesis extends previous work on peak-load pricing in two respects. First, it links the determination of peak-load prices to a mathematical programming model for long-range capacity planning. Second, it uses the probabilistic version of the capacity planning model, presented in Chapter 4, so that the effects of system reliability on prices can be considered.

The linkage of peak-load pricing to a mathematical model for long range capacity planning provides a way of calculating marginal costs for generating systems with many types of plants. A major shortcoming of some of the previous work in peak-load pricing is that planning (and optimizing) over a multi-year horizon is not considered. Instead, the capital costs of the generating plants are put on an annualized basis, and the optimal plant capacities and prices are computed for each year separately. However, the process of annualizing capital costs does not necessarily lead to a correct allocation of the costs over the life of the plant. In a multi-year capacity planning model, on the other hand, the plant's capital costs are automatically allocated to each year of its life on the basis of the <u>shadow value</u> of having its capacity available in that year. For example, in the decomposition models discussed above, the shadow prices λ_{jv}^{t} give the economic value of each plant in each year t. The optimality condition (4.15) written as

$$C_{jv} - \sum_{t=1}^{T} \lambda_{jv}^{t} = 0 \quad \text{when} \quad X_{jv} > 0$$

requires that the discounted sum of these benefits derived from operating a plant over the planning horizon must equal the initial cost of building the plant in order for the plant to be built. Thus, these shadow prices serve to allocate the capital costs of the plants according to economic value in each year.[2] In addition, the marginal costs derived from the mathematical programming model are forward-looking, rather than embedded, costs, and therefore reflect the opportunity costs of supplying electricity.

Using probabilistic models for peak-load pricing gives some conclusions which are different than those reached in the traditional models. Traditional work reached the conclusion that since peak users are the ones who determine the need for additional capacity, they (and only they) should be charged the marginal capacity costs. More recent work, however, has shown that when the system design goal is based on meeting all loads with a specified level of reliability, all users should be charged a portion of the marginal capacity cost, in proportion to their contribution to the risk of loss of load[3]. Furthermore, since all plants contribute to meeting the reliability goal, all plants contribute to the marginal capacity cost.

This part is divided into three chapters. Chapter 10 discusses the derivation of marginal cost information from the probabilistic capacity planning model. First, a technique for calculating the marginal costs attributable to different components of the load duration curve is presented. This technique is based on calculating the loss-of-load probability associated with each component using a recursive procedure similar to probabilistic simulation. Second, these marginal costs are derived from the probabilistic planning model, so that they can be calculated as a by-product of the capacity expansion optimization. It is pointed out that because the functions involved are not differentiable at all points, these marginal costs do not correspond to the usual definition. Instead, they represent a linear support, or subgradient, of the cost function.

Chapter 11 discusses, in more rigorous detail, the problems associated with the non-differentiability of the cost function in Chapter 10. Techniques from the theory of non-differentiable optimization are used to show that solving the capacity planning problem by generalized Benders' decomposition produces subgradients of the cost function, and the marginal cost expressions given in Chapter 10 can be derived from these subgradients.

Finally, Chapter 12 discusses the use of the capacity planning model within a peak-load pricing model. The model proposed brings in the demand for electricity and finds an equilibrium between supply and demand. A decomposition procedure is proposed for solving this equilibrium problem, in which the capacity planning model is used as a subproblem to calculate supply cost. The subgradients derived in Chapter 11 are used to carry cost information from the subproblems into the equilibrium problem.

CHAPTER 10
MARGINAL COST CALCULATIONS USING THE CAPACITY PLANNING MODEL

A. Introduction

The capacity planning models discussed in previous chapters calculate the minimum cost of satisfying a given demand for electricity. One can view these models, therefore, as defining cost as a function of demand. For pricing purposes, it is necessary to know how the cost of supplying electricity varies when demand changes. Of particular interest are the marginal cost changes which result from marginally varying a given demand. An advantage of using a mathematical programming model to define the cost function is that such marginal cost information is readily available in the shadow prices associated with the optimal solution.

This chapter discusses the use of the probabilistic capacity planning model, presented in Chapter 4, in the calculation of marginal costs. The calculation has three aspects. First, since the demand for electricity in this model is specified in the form of a load duration curve, the notion of a marginal change in demand must be precisely defined. The load duration curve can be viewed as consisting of a set of discrete components, each representing the demand in a single time interval (usually, one hour), arranged in order of decreasing magnitude. A marginal change in the load duration curve can then be represented as a marginal change in one of these components.

Second, the costs of supplying electricity must be allocated among the different components of the load duration curve. It will be shown that the correct allocation is based on the loss-of-load probability associated with each component. A recursive procedure for computing these probabilities, related to the probabilistic simulation procedure, will be presented.

Third, the marginal costs themselves must be derived from the capacity planning model. This is accomplished by reformulating the model presented in Chapter 4 to explicitly include the contribution to cost made by each component of the load curve. The marginal costs can then be derived from the optimality conditions for the problem. This derivation is somewhat complicated by the non-differentiability of the functions involved, a topic which is discussed in greater detail in Chapter 11.

In Section B of this chapter, the recursive procedure is presented for calculating the loss-of-load probability associated with different load levels. This result is used in Section C to derive expressions for the total and marginal cost contributions attributable to different load levels. In Section D, the probabilistic capacity planning model is reformulated, on the basis of the expressions derived in Section C, to include explicitly the contribution to cost of each component of the load curve. Expressions for marginal cost are derived and some of the difficulties due to non-differentiability are pointed out.

The method of allocating costs in a probabilistic environment presented in this chapter is based on the work of Vardi, et al[1]. They have made the important observation that, in a probabilistic environment, since each hour of the year contributes to the system loss-of-load probability, each hour should be allocated part of the marginal capacity cost for achieving the system-wide reliability goal. They have presented a method for calculating the appropriate contribution of each hour to these costs based on its contribution to the loss-of-load probability. However, since their method requires enumeration of the outage states, the calculation can be costly for systems with many plants. The reformulation presented in Section B of this chapter uses a recursive computation which would appear to be more efficient.

The paper of Vardi, et al., assumes that the marginal costs, particularly the marginal capacity cost, are known. Several authors have discussed the use of mathematical programming models for calculating marginal cost. Scherer[2] uses a mixed integer programming model to determine the optimal capacities for the various types of plants, and he calculates the marginal costs from the shadow prices associated with the optimal solution. However, his model does not calculate a capacity expansion plan over a multi-year planning horizon. Instead it uses annualized capital costs to calculate the optimal plant capacities and prices for each year individually. The work of Telson[3], in which he calculates the costs and benefits of increased generating system reliability, is based on a mathematical programming capacity expansion model, the GEM model, which uses probabilistic simulation to calculate system reliability. However, GEM does not use shadow prices to link the probabilistic simulation subproblem to the capacity planning linear program, and Telson himself does not make use of the LP shadow prices in his analysis. The model presented in this chapter therefore unifies several approaches to calculating marginal costs for supplying electricity, by using the shadow prices from a mathematical programming capacity planning model and by using an allocation of costs based on probabilistic criteria.

B. Hour-by-Hour Allocation of Loss-of-Load Probability

This section presents a method for calculating the loss-of-load probability (LOLP) attributable to demand in a given component (or hour) of the load duration curve. The argument used in the derivation is a recursive one similar to that used to derive the probabilistic simulation in Chapter 4. The argument is simplest when pursued in the domain of load rather than in the domain of time. However, the section concludes with a comparison of this method with that of Vardi, et al., which works in the time domain.

Define the function $g_i(Q)$ to be the probability of loss of load faced by the $i^{\underline{th}}$ plant in the merit order when the load is Q. Since there are no plants below the first plant in the merit order, when $i = 1$, loss of load is certain if the load is positive and impossible if the load is not positive. Hence define

$$g_1(Q) = \begin{cases} 1 & \text{if } Q > 0 \\ 0 & \text{if } Q \leq 0 \end{cases} \tag{10.1}$$

To find the loss-of-load probability (LOLP) faced by the other plants, a conditional probability argument is used, similar to the argument used in probabilistic simulation. Given that plant i does not operate, the LOLP faced by plant i+1 is the same as that faced by plant i, namely $g_i(Q)$. On the other hand, if plant i operates at level Y^i, then at least Y^i units of the load will be served. The probability that the plants below i will fail to serve the remaining $Q-Y^i$ units of load is $g_i(Q-Y^i)$ (note that $g_i(Q) = 0$ if $Q \leq 0$). Weighting each event by its probability gives

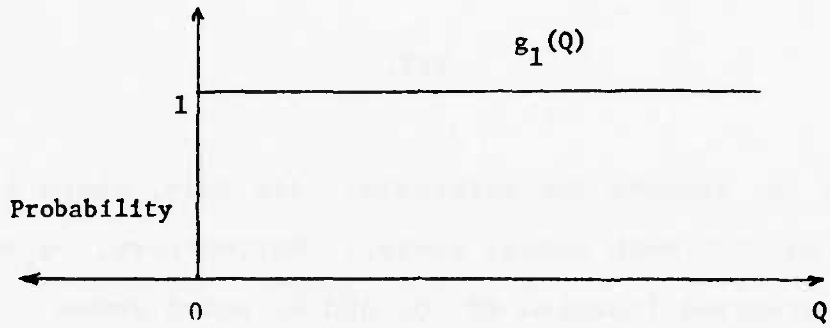
$$g_{i+i}(Q) = p_i g_i(Q-Y^i) + q_i g_i(Q) \tag{10.2}$$

Equations (10.1) and (10.2) serve to define the family $g_i(Q)$ recursively. Note that, once all of the plants have been loaded, the loss-of-load probability of the system when the load is Q is

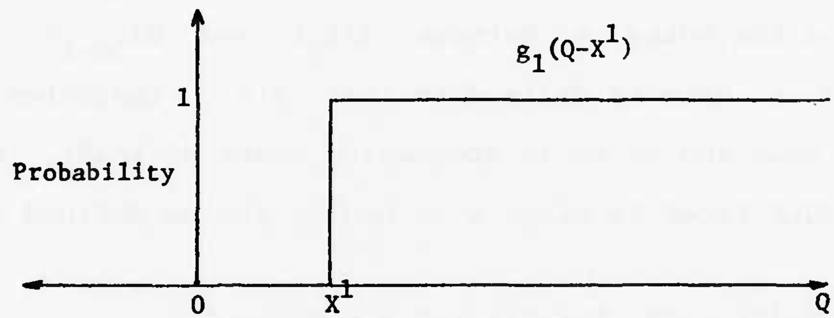
$$g_{I+1}(Q).$$

Several properties of the g_i functions follow immediately from the recursive definition. They are step functions; that is, the probability of loss-of-load is constant on intervals of load

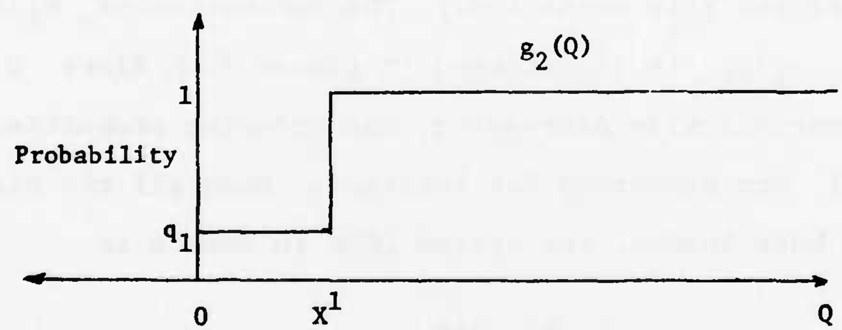
$$g_i(Q) = \{\psi_\sigma \text{ for } \xi_{\sigma-1} < Q \leq \xi_\sigma\}$$



1. If there is only one plant, and it fails, loss-of-load is certain for psitive loads. This event occurs with probability q_1.



2. When the first plant operates, loss-of-load is certain only for loads greater than capacity, X^1. This event occurs with probability p_1.



3. The loss-of-load probability with one plant is the sum of these two curves weighted by their probabilities.

Figure 10.1

Derivation of the Probability function g_1

where σ indexes the intervals. (In fact, there is one interval for each outage state.) Furthermore, $g_i(Q)$ is an increasing function of Q, and as noted above $g_i(Q) = 0$ for $Q \leq 0$.

By the definition of the load duration function $G(Q)$, the load Q lies in the interval between $\xi_{\sigma-1}$ and ξ_σ during the hours s between $G(\xi_\sigma)$ and $G(\xi_{\sigma-1})$ (the index s here is defined so that $s = 0$ indicates the peak hour and so on in decreasing order of load). Hence, the LOLP faced by plant i in hour s can be defined as

$$h_i(s) = \{\psi_\sigma \text{ for } G(\xi_\sigma) \leq s < G(\xi_{\sigma-1})\}$$

(Actually, since the load duration function $G(Q)$ may not be one-to-one, $h_i(s)$ must be defined as the maximum ψ_σ satisfying this condition.) The derivation of $h_i(s)$ from $g_i(Q)$ is illustrated in Figure 10.2. Since $G(Q)$ is monotonically decreasing, the ordering properties of $g_i(Q)$ are preserved but inverted. When all the plants have been loaded, the system LOLP in hour s is

$$h_{I+1}(s)$$

In their paper[4], Vardi, *et al.*, present another method for calculating this contribution of hour s to

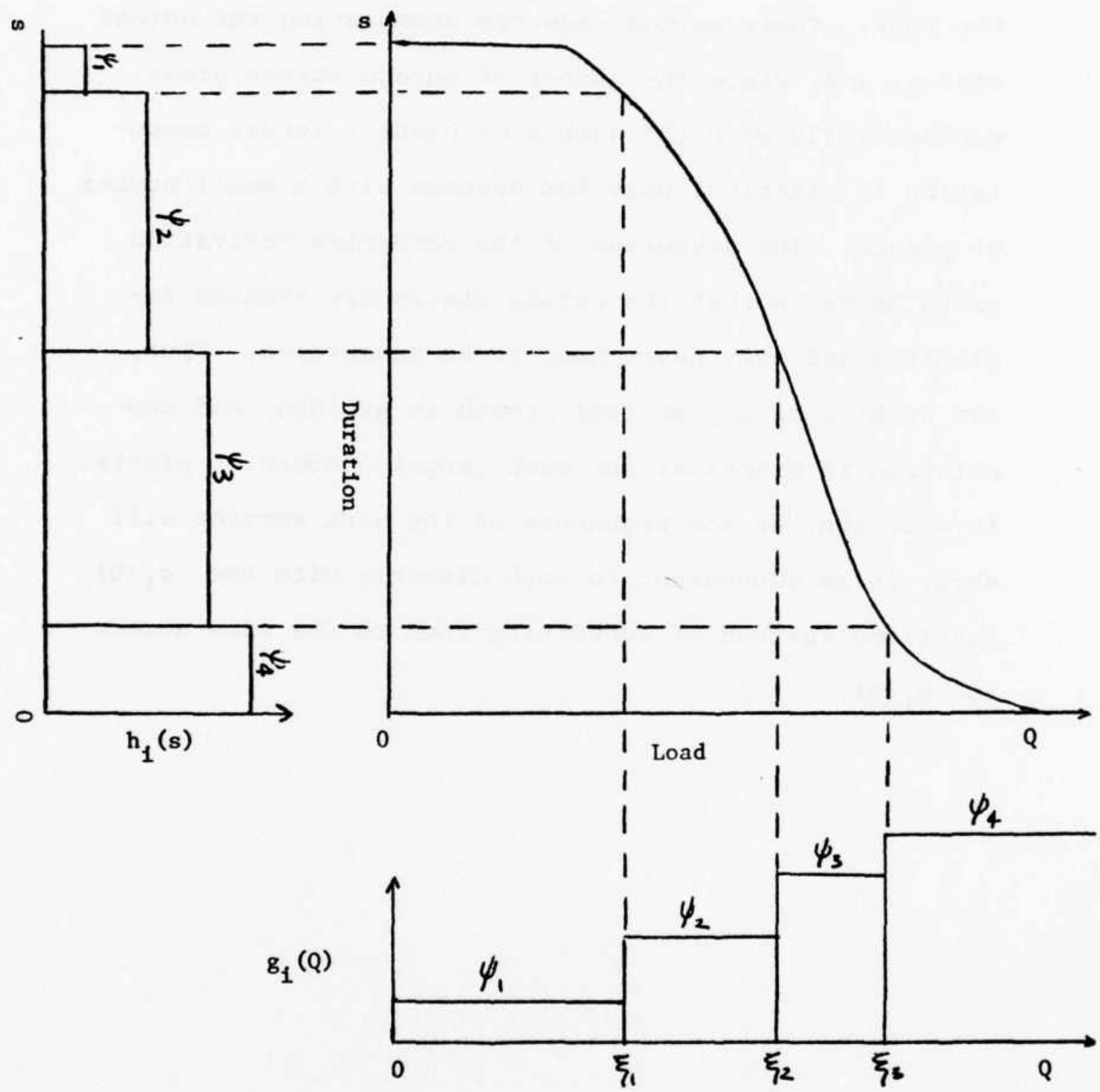


Figure 10.2

Derivation of the Probability Function h_i

the LOLP. Their method requires enumerating the outage states, and, since the number of outage states grows exponentially with the number of plants, actual computation is practical only for systems with a small number of plants. The advantage of the recursive derivation given above is that the outage states are treated implicitly and they never need to be enumerated. Thus, the problem of exponential growth is avoided, and computation is practical for much larger numbers of plants. In addition, as the arguments of the next section will show, it is convenient to work directly with the $g_i(Q)$ functions instead of converting them to the time domain as $h_i(s)$.

C. Determination of Hourly Contributions to Cost and Unserved Energy

When plant i is the marginal plant, the marginal operating cost is F^i. Given that plant i is operating, the probability that it is the marginal plant is the probability that the load falls above the point at which plant i is loaded and below its utilization level. More precisely, let $\tilde{U}^{i-1}$ be the (random) loading point of plant i, the total output of all the available plants below i in the merit order (clearly $0 \leq \tilde{U}^{i-1} \leq U^{i-1}$). Then given that plant i operates, the probability that it is marginal when the load is Q is the probability that

$$\tilde{U}^{i-1} < Q \leq \tilde{U}^{i-1} + Y^i$$

However, the LOLP faced by plant i when the load is Q is just the probability that

$$\tilde{U}^{i-1} < Q.$$

Hence, given that plant i operates, the probability that it is the marginal plant is

$$g_i(Q) - g_i(Q-Y^i)$$

Since the probability that plant i operates is p_i, the unconditional probability that plant i is the marginal plant when the load is Q is

$$p_i[g_i(Q)-g_i(Q-Y^i)]$$

which is equal to

$$g_i(Q) + g_{i+1}(Q)$$

Thus the expected marginal operating cost when the load is Q is

$$\sum_{i=1}^{I} F^i p_i[g_i(Q)-g_i(Q-Y^i)] \tag{10.3}$$

Note also that the probability that the load exceeds the available system capacity, in which case no plant is marginal, is just the LOLP when the load is Q,

$$g_{I+1}(Q)$$

If the expresssion (10.3) is the marginal operating cost when the load is Q, then its integral should be the total operating cost. The proof of this assertion, which follows, leads to a useful correspondence with the expression for cost in the probabilistic model derived previously, in Chapter 4.

Define a special load duration function, the step function

$$G(Q,Q^*) = \begin{cases} 1 & \text{if} \quad Q < Q^* \\ 0 & \text{if} \quad Q \geq Q^* \end{cases} \tag{10.4}$$

where Q^* is the peak load. For this function, the duration of all loads less than or equal to the peak is the entire period. The equivalent load duration curves $G_i(Q,Q^*)$ are defined by the probabilistic simulation recursion, as usual. An important property of the step function is that

$$G(Q-Y,Q^*) = G(Q,Q^*+Y) \tag{10.5}$$

and by induction, this property holds for all of the equivalent load duration curves $G_i(Q,Q^*)$ as well.

The following property will be useful:

First Proposition:

$$\int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ = \int_{Q^*-Y}^{Q^*} g_i(Q)dQ \tag{10.5}$$

Proof:

$$\int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ$$

$$= \int_0^{Q^*} g_i(Q)dQ - \int_0^{Q^*} g_i(Q-Y)dQ$$

$$= \int_0^{Q^*} g_i(Q)dQ - \int_0^{Q^*-Y} g_i(Q)dQ$$

(since $g_i(Q) = 0$ for $Q \leq 0$)

$$= \int_{Q^*-Y}^{Q^*} g_i(Q)dQ \quad ||$$

It is desired to show that

$$\sum_{i=1}^{I} F^i p_i \int_0^{Q^*} [g_i(Q)-g_i(Q-Y^i)]dQ$$

$$= \sum_{i=1}^{I} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q,Q^*)dQ \tag{10.7}$$

where the latter expression is the operating cost as derived in Chapter 4, for the step load duration curve. This correspondence is established by the following proposition

Second Proposition: For all Q^* and $Y \geq 0$,

$$\int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ = \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q,Q^*)dQ \qquad (10.8)$$

<u>Proof</u>: The proof is by induction. When $i = 1$,

$$g_i(Q) = \begin{cases} 1 & \text{if} \quad Q > 0 \\ 0 & \text{if} \quad Q \leq 0 \end{cases}$$

Hence $\int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ$

$$= \int_{Q^*-Y}^{Q^*} g_1(Q)dQ = \begin{cases} Q^* & \text{if} \quad Q^* \leq Y \\ Y & \text{if} \quad Q^* > Y \end{cases}$$

Also

$$\int_0^{Y} G_1(Q,Q^*)dQ = \begin{cases} Q^* & \text{if} \quad Q^* \leq Y \\ Y & \text{if} \quad Q^* > Y \end{cases}$$

Hence equality holds when $i = 1$. Now assume that (10.8) holds for some i. Then

$$\int_0^{Q^*} [g_{i+1}(Q)-g_{i+1}(Q-Y)]dQ$$

$$= p_i \int_0^{Q^*} [g_i(Q-Y^i)-g_i(Q-Y^i-Y)]dQ$$

$$+ q_i \int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ \qquad \text{(by (10.2))}$$

$$= p_i \int_0^{Q^*-Y^i} [g_i(Q)-g_i(Q-Y)]dQ$$

$$+ q_i \int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ$$

since $g_i(Q) = 0$ for $Q \leq 0$. Now, by the inductive assumption

$$\int_0^{Q^*} [g_i(Q)-g_i(Q-Y)]dQ = \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q,Q^*)dQ$$

and

$$\int_0^{Q^*-Y^i} [g_i(Q)-g_i(Q-Y)]dQ = \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q,Q^*-Y^i)dQ$$

$$= \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q+Y^i,Q^*)dQ$$

by property (10.5). Therefore

$$\int_0^{Q^*} [g_{i+1}(Q)-g_{i+1}(Q-Y)]dQ$$

$$= p_i \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q+Y^i,Q^*)dQ + q_i \int_{U^{i-1}}^{U^{i-1}+Y} G_i(Q,Q^*)dQ$$

$$= p_i \int_{U^i}^{U^i+Y} G_i(Q,Q^*)dQ + q_i \int_{U^i}^{U^i+Y} G_i(Q-Y^i,Q^*)dQ$$

(since $U^i - U^{i-1} = Y^i$)

$$= \int_{U^i}^{U^i+Y} G_{i+1}(Q,Q^*)dQ,$$ which proves the proposition. ||

An extension of the previous proposition gives an alternative expression for expected unserved energy:

<u>Third Proposition</u>:

$$\int_0^{Q^*} g_{I+1}(Q)dQ = \int_{U^I}^{\infty} G_{I+1}(Q,Q^*)dQ \tag{10.9}$$

<u>Proof</u>:

$$\int_{Q^*-Y}^{Q^*} g_{I+1}(Q)dQ = \int_0^{Q^*} [g_{I+1}(Q)-g_{I+1}(Q-Y)]dQ$$

(by the first proposition (10.6))

$$= \int_{U^I}^{U^I+Y} G_{I+1}(Q,Q^*)dQ$$

(by the s nd proposition (10.8)).

Letting $Y \to \infty$ gives

$$\int_0^{Q^*} g_{I+1}(Q)\,dQ = \int_{U^I}^{\infty} U_{I+1}(Q,Q^*)\,dQ$$

since $g_{I+1}(Q) = 0$ for $Q \leq 0$. ||

It should be noted that because the functions $g_i(Q)$ are step functions, the integrals in (10.7) and (10.9) are not differentiable at all points. Points of discontinuity of g_i correspond to loads for which unique marginal costs cannot be defined. The implications of the non-differentiability of these cost functions will be discussed in the next section and in depth in Chapter 11.

D. <u>A Model for Determinating Supply Cost and Capacity Planning</u>

The capacity expansion planning model developed in Chapter 4 can be reformulated, using the cost expressions derived in the previous chapter. This new formulation can be used to derive the marginal costs associated with the demand components in different hours.

Consider first the operating subproblem in a single year, with plant capacities and demand fixed. The load duration curve can be approximated by discrete components. Let Q_s be the demand in hour s, where as before, the index s is defined so that

$$Q_1 \geq Q_2 \geq \dots \geq Q_S$$

As in the previous section, define the step function

$$G(Q,Q_s) = \begin{cases} 1 & \text{if} \quad Q \leq Q_s \\ 0 & \text{if} \quad Q > Q_s \end{cases}$$

Then the load duration curve for the year can be approximated by

$$G(Q) = \sum_{s=1}^{S} G(Q,Q_s) = \text{number of hours in which the load}$$

exceeds Q.

(Note that this approximation can actually be made arbitrarily close to any load duration curve by letting the time increment corresponding to demand Q_s be small enough and the number of increments S be very large. Hours are used here as the increment for convenience.) The equivalent load duration curve faced by each plant i can be defined using the probabilistic simulation recursion as before and

$$G_i(Q) = \sum_{s=1}^{S} G_i(Q,Q_s).$$

Using this load duration curve, the usual operating subproblem can be written, as in Chapter 4,

$$\text{minimize} \quad \sum_{i=1}^{I} \sum_{s=1}^{S} F^i p_i \int_{U^{i-1}}^{U^i} G_i(Q,Q_s)\,dQ$$

$$\text{subject to} \quad \sum_{s=1}^{S} \int_{U^I}^{\infty} G_{I+1}(Q,Q_s)\,dQ \leq \varepsilon$$

$$0 \leq Y^i \leq X^i$$

$$\text{where} \quad U^i - U^{i-1} = Y^i \qquad i = 1,\ldots,I$$

$$\text{with} \quad U^0 = 0.$$

This model can be rewritten using (10.7) and (10.9) as

$$\text{minimize} \quad \sum_{i=1}^{I} \sum_{s=1}^{S} F^i \int_0^{Q_s} [g_i(Q)-g_{i+1}(Q)]dQ \qquad (10.10)$$

$$\text{subject to} \quad \sum_{s=1}^{S} \int_0^{Q_s} g_{i+1}(Q)dQ \leq \varepsilon \qquad (10.11)$$

$$0 \leq Y^i \leq X^i \qquad (10.12)$$

Using this operating subproblem, the capacity planning problem can be written

$$TC = \underset{\underline{X},\underline{Y}_1,\ldots,\underline{Y}_T}{\text{minimum}} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t,\underline{Q}_t) \qquad (10.13)$$

$$\text{subject to} \quad EG_t(\underline{Y}_t,\underline{Q}_t) \leq \varepsilon_t \qquad (10.14)$$

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

$$0 \leq \underline{Y}_t \leq \delta_t\underline{X} \qquad (10.15)$$

where the load duration curve in period t is represented by the vector of hourly components $\underline{Q}_t$, the objective function (10.10) is the operating cost function $EF_t(\underline{Y}_t,\underline{Q}_t)$ and the unserved energy in (10.11) is represented by the function $EG_t(\underline{Y}_t,\underline{Q}_t)$.

This model (10.13) - (10.15) defines a function $TC(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T)$, the minimum (long-run) cost of meeting demand $\underline{Q}_t$ with reliability ε_t in each year t of the planning horizon. The reliability level ε_t is included as a parameter of demand in order to allow calculation of the marginal cost of increased reliability. When the cost function TC is used in an equilibrium model to determine prices, as will be discussed in Chapter 12, the equilibrium reliability levels, as well as the equilibrium supply quantities, can be determined. This approach is equivalent to the inclusion of rationing costs to determine reliability, as was mentioned in Chapter 4, and thus it represents a resolution of the two approaches - cost and constraint - to reliability.

The purpose of defining the cost function TC is to be able to get some notion of how costs vary with cahnges in demand and reliability, the marginal costs. However, the function TC is not necessarily differentiable, so that marginal costs, in the usual sense, cannot be precisely defined. Nevertheless, the function TC is convex (as will be shown in Chapter 11), and therefore, it possesses directional derivatives in all directions. Furthermore, at each point it possesses linear support functionals, or <u>subgradients</u>, which are used to compute the directional derivatives, and which can be used (with care) to stand in for the gradient. Consider first a naive approach to finding marginal costs.

Suppose an optimal solution to (10.13) - (10.15) has been found. Intuitivaly, the shadow price π_t on the unserved energy constraint (10.14) represents the marginal capacity cost of the system, since the model would build just enough capacity to meet the reliability standard ε_t. Thus, the cost of slightly changing that standard is just the marginal cost of additional capacity. The marginal contribution of demand in hour s, Q_{ts}, to the unserved energy can be written

$$\frac{\partial EG_t}{\partial Q_{ts}}$$

evaluated at the optimal solution (pretending that this derivative exists). The marginal contribution of this demand to operating costs can be written

$$\frac{\partial EF_t}{\partial Q_{ts}} \, .$$

Thus, the marginal contribution to cost of this demand is

$$MC(Q_{ts}) = \frac{\partial EF_t}{\partial Q_{ts}} + \pi_t \frac{\partial EG_t}{\partial Q_{ts}} \qquad (10.16)$$

Formally differentiating (10.16) gives

$$MC(Q_{ts}) = \sum_{i=1}^{I_t} F^{it}[g_{it}(Q_{ts}) - g_{i+1,t}(Q_{ts})]$$

$$+ \pi_t g_{I_t+1}(Q_{ts}) \qquad (10.17)$$

which is in the form of the marginal costs given in Section B. Although the functions EF_t and EG_t are actually not differentiable, (10.17) gives a valid expression for the subgradients of the cost function TC, as will be shown in Chapter 11.

Now, in solving the problem (10.13) - (10.15) by generalized Benders' decomposition, π_t will not be directly available, since what will actually be solved is the master problem

$$\underset{Z,\underline{X}}{\text{minimize}}\ Z \qquad (10.18)$$

subject to

$$Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} [EF_t(\underline{Y}_t^k,\underline{Q}_t) + \underline{\lambda}^k \delta_t(\underline{X}^k-\underline{X})] \qquad (10.19)$$

$$k = 1,\ldots,K$$

$$\sum_{t\epsilon\Gamma^k} [EG_t(\underline{Y}_t^k,\underline{Q}_t) + \underline{\mu}_t^k \delta_t(\underline{X}^k-\underline{X})-\epsilon_t] \leq 0 \qquad (10.20)$$

$$\underline{X} \geq \underline{0}$$

which is a linear program. However, as will be shown in the next chapter, the proper subgradient can be found from the shadow prices associated with this master problem. Let θ^k be the dual multiplier associated with the $k^{\underline{th}}$ cost constraint (10.17) and ν^k be the dual multiplier associated with the $k^{\underline{th}}$ feasibility constraint (10.18). Then it will be shown in the next chapter that a subgradient can be computed which has components

$$MC(Q_{ts}) = \sum_{k=1}^{K} \theta^k [\sum_{i=1}^{I_t} F^{it}(g_i(Q_{ts};\underline{Y}_t^k) - g_{i+1}(Q_{ts};\underline{Y}_t^k))]$$

$$+ \sum_{k=1}^{K} \nu^k g_{I_t+1}(Q_{ts};\underline{Y}_t^k) \qquad (10.21)$$

and

$$MC(\varepsilon_t) = \sum_{k=1}^{K} \{\theta^k \pi_t^k + \nu^k\} \qquad (10.22)$$

where π_t^k is the shadow price on unserved energy computed in the subproblem for period t at iteration k. The notation $g_i(Q;\underline{Y}_t^k)$ indicates that the functions g_i are computed using the optimal solution $\underline{Y}$ to the subproblem in period t at iteration k.

It should be noted that the subgradient thus computed is not unique and therefore does not correspond to marginal cost in the usual sense. However, in Chapter 12 it will be shown that such a subgradient can be used in place of marginal cost in computing peak-load prices.

CHAPTER 11

SOME TECHNICAL ISSUES

A. Introduction

This chapter discusses more rigorously some of the results presented in the previous chapter. In the previous chapter, formulas were given for the marginal contribution to cost associated with demand in each hour and with the reliability (unserved energy) standard in each period. It was noted that because of the non-differentiability of the functions involved, marginal costs calculated are not necessarily unique and so do not correspond to the usual notion of marginal cost. Instead, they correspond to components of the subgradient of the cost function. This chapter presents a rigorous derivation of these formulas.

Section B derives some results on subgradients for general mathematical programs. The discussion considers a function $v(\underline{y})$ defined as the optimal value of a mathematical program which depends on a vector of parameters $\underline{y}$. It is shown that subgradients of this function can be derived from the shadow prices associated with the optimization. Two cases are discussed. First, when the objective function and constraints of the program are differentiable, these shadow prices are derived from the usual Kuhn-Tucker optimality conditions. Second, when the objective function and constraints are not differentiable, they are approximated by their subgradients. An algorithm similar to the generalized Benders' algorithm discussed in Part Two is used to generate the pieces of this approximation, and the problem is solved by solving a sequence of linear programs which contain the subgradient approximation. Then the required shadow prices and the subgradient of v can be derived from the dual solution to the optimal linear program.

In Section C, these general results are applied to calculating a subgradient of the cost function TC defined by the capacity expansion planning model. It is shown that solving the operating subproblems to generate a Benders' cut for the master problem is essentially the same as determining a subgradient approximation to the objective function and constraints of the capacity planning model. Furthermore, these subgradients correspond to the optimal shadow prices of the subproblem. These shadow prices are calculated by the formulas derived in Chapter 9 which, though originally derived from the

Kuhn-Tucker conditions, remain valid in the non-differentiable case. The elements of the subgradients corresponding to the components of the load duration curve and to the reliability level are derived from the marginal cost formulas given in Chapter 10. The generalized Benders' master problem then corresponds to the subgradient approximation problem of Section B. Finally, it is shown how to compute a subgradient of the cost function TC using the shadow prices in the optimal master problem and the subgradients with respect to demand and reliability generated by the subproblems.

The derivations in this chapter are based strongly on the works of Geoffrion[1], Hogan[2], and Shapiro[3].

B. Some General Results

Consider a general parametric mathematical program

$$\begin{aligned} v(\hat{\underline{y}}) &= \underset{x}{\text{minimum}}\ f(\underline{x},\hat{\underline{y}}) \\ &\text{subject to } g(\underline{x},\hat{\underline{y}}) \leq 0 \end{aligned} \tag{11.1}$$

where $\underline{x}$ is an n-dimensional vector, $\hat{\underline{y}}$ is k-dimensional, f maps $R^n \times R^k$ into the real numbers and g maps $R^n \times R^k$ into R^m. This program can be written

$$\begin{aligned} v(\hat{y}) &= \underset{x,w}{\text{minimum}}\ f(\underline{x},\underline{w}) \\ &\text{subject to } g(\underline{x},\underline{w}) \leq 0 \\ &\underline{w} = \hat{\underline{y}} \end{aligned} \tag{11.2}$$

where $\underline{w}$ is a vector of auxiliary variables. Clearly, this auxiliary problem has the same optimal solution as the original. Let $\underline{\lambda}$ be a vector of dual multipliers associated with the original constraints $g(\underline{x},\underline{w}) \leq 0$ and let $\underline{\gamma}$ be the vector of multipliers associated with the auxiliary constraints $\underline{w} = \hat{\underline{y}}$. Assume that for each $\hat{y}$ for which the program has a feasible solution, the optimal solution $(\hat{\underline{x}},\hat{\underline{w}})$ and the optimal multipliers $(\hat{\underline{\lambda}},\hat{\underline{\gamma}})$ satisfy the global optimality conditions (Theorem 2 in the Introduction to Part Two).

Even if it is assumed that f and g are differentiable everywhere, it still cannot be guaranteed that $v(\underline{y})$ is differentiable everywhere. A simple example is provided by the linear program in which $f(\underline{x},\underline{y}) = \underline{c}'\underline{x}$ and $g(\underline{x},\underline{y}) = \underline{y}-A\underline{x}$. The values of $\underline{y}$ at which the optimal basis changes are points where the function $v(\underline{y})$ is not differentiable. However, $v(\underline{y})$ does possess linear support functionals, or <u>subgradients</u>, and directional derivatives in all directions. A subgradient at a point $\hat{\underline{y}}$ is a vector $\hat{\underline{\gamma}}$ such that for any other point $\underline{y}$

$$v(\underline{y}) \geq v(\hat{\underline{y}}) + \hat{\underline{\gamma}}(\underline{y}-\hat{\underline{y}}) \tag{11.3}$$

If v is differentiable at $\hat{\underline{y}}$ then $\hat{\underline{\gamma}}$ is unique and equal to the gradient of v. However, if v is not differentiable at $\hat{\underline{y}}$, then $\hat{\underline{\gamma}}$ is not unique; the set of all subgradients at $\hat{\underline{y}}$ is called the <u>subdifferential</u>, denoted $\partial v(\hat{\underline{y}})$. The directional derivative of v at $\hat{\underline{y}}$ in the direction $\Delta\underline{y}$ can be found by solving the mathematical program

$$\begin{aligned} \dot{v}(\hat{\underline{y}}|\Delta\underline{y}) = \text{maximum } & \underline{\gamma}\cdot\Delta\underline{y} \\ \text{subject to } & \underline{\gamma} \in \partial v(\hat{\underline{y}}) \end{aligned} \tag{11.4}$$

It is not difficult to show that there is a relationship between the dual multipliers of (11.2) and the subgradients of v.

<u>First Proposition</u>: For a given point $\hat{\underline{y}}$, if $\hat{\underline{\gamma}}$ is the optimal multiplier vector on the auxiliary constraints in (11.2) and the primal and dual solutions satisfy the global optimality conditions, then $\hat{\underline{\gamma}}$ is a subgradient of v at $\hat{\underline{y}}$.

<u>Proof</u>: Let $\underline{y}$ be any other point. Let $(\hat{\underline{x}},\hat{\underline{w}})$ be optimal in (11.2). Then

$$v(\hat{\underline{y}}) = f(\hat{\underline{x}},\hat{\underline{w}})$$

By the global optimality conditions

$$v(\hat{\underline{y}}) = \underset{\underline{x},\underline{w}}{\text{minimum}}\ f(\underline{x},\underline{w}) + \hat{\underline{\lambda}}g(\underline{x},\underline{w}) + \hat{\underline{\gamma}}(\hat{\underline{y}}-\underline{w})$$

where $\hat{\underline{\lambda}}$ and $\hat{\underline{\gamma}}$ are the optimal dual multipliers in (11.2). By the Strong Duality Property (Theorem 3 of the Introduction to Part Two),

$$v(\underline{y}) = \underset{\underline{\lambda} \geq 0,\underline{\gamma}}{\text{maximum}}\ \underset{\underline{x},\underline{w}}{\text{minimum}}\ f(\underline{x},\underline{w}) + \underline{\lambda}g(\underline{w},\underline{w}) + \gamma(\underline{y}-\underline{w})$$

$$\geq \underset{\underline{x},\underline{w}}{\text{minimum}}\ f(\underline{x},\underline{w}) + \hat{\underline{\lambda}}g(\underline{x},\underline{w}) + \hat{\underline{\gamma}}(\underline{y}-\underline{w})$$

$$= f(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\lambda}}g(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\gamma}}(\underline{y}-\hat{\underline{y}}+\hat{\underline{y}}-\hat{\underline{w}})$$

$$= f(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\gamma}}(\underline{y}-\hat{\underline{y}})$$

(by complementary slcakness)

$$= v(\hat{y}) + \hat{\gamma}(y-\hat{y}). \quad ||$$

Notice that this proof does not use any convexity properties. However, subgradients are associated with convex functions, and therefore it might be suspected that v is convex.

Corollary: $v(y)$ is a convex function.

Proof: Let y_1 and y_2 be any two points and let

$$\hat{y} = \alpha y_1 + (1-\alpha)y_2 \quad \text{where} \quad 0 \leq \alpha \leq 1.$$

Let $\hat{\gamma}$ be a subgradient of v at $\hat{y}$. Then

$$v(y_1) \geq v(\hat{y}) + \hat{\gamma}(y_1-\hat{y}) = v(\hat{y}) + \hat{\gamma}(1-\alpha)(y_1-y_2)$$

and

$$v(y_2) \geq v(\hat{y}) + \hat{\gamma}(y_2-\hat{y}) = v(\hat{y}) + \hat{\gamma}(-\alpha)(y_1-y_2).$$

Hence

$$\alpha v(y_1) + (1-\alpha)v(y_2) \geq v(\hat{y})$$

which proves that $v(y)$ is convex. ||

Because of this First Proposition, it is sufficient to find an optimal dual solution to (11.2) in order to find a subgradient. However, to find directional derivatives it is necessary to construct the subdifferential, the set of all subgradients. This task is somewhat more difficult.

When f and g are differentiable and convex, Geoffrion[4] has shown how to construct the directional derivatives. In this case the Kuhn-Tucker conditions characterize the optimal dual solutions. The Kuhn-Tucker conditions for (11.2) are

i) Stationarity

$$\frac{\partial f}{\partial \underline{x}} + \underline{\lambda}\, \frac{\partial g}{\partial \underline{x}} = 0$$

$$\frac{\partial f}{\partial \underline{w}} + \underline{\lambda}\, \frac{\partial g}{\partial \underline{w}} - \underline{\gamma} = 0$$

ii) Complementary Slackness

$$\underline{\lambda} g(x,w) = 0$$

(the gradients $\frac{\partial f}{\partial \underline{x}}$, $\frac{\partial g}{\partial \underline{x}}$, $\frac{\partial f}{\partial \underline{w}}$ and $\frac{\partial g}{\partial \underline{w}}$ are evaluated at the optimal solution to the problem). Then, the mathematical program for the directional derivative (11.4) becomes

$$\dot{v}(\underline{\gamma}|\Delta\underline{\gamma}) = \underset{\underline{\lambda}\geq 0}{\text{maximum}} \left(\frac{\partial f}{\partial \underline{w}} + \underline{\lambda}\frac{\partial g}{\partial \underline{w}}\right)\cdot\Delta\underline{\gamma}$$

$$\text{subject to} \quad \frac{\partial f}{\partial \underline{x}} + \underline{\lambda}\, \frac{\partial g}{\partial \underline{x}} = 0 \tag{11.5}$$

$$\lambda_k = 0 \quad \text{if} \quad g_k(\underline{x},\underline{w}) < 0$$

where λ_k is the $k^{\underline{th}}$ component of $\underline{\lambda}$, and all gradients are evaluated at the optimal solution to (11.2). Notice that this problem is a linear program.

When f and g are not differentiable (but are still convex), the situation is more complicated because the Kuhn-Tucker conditions no longer characterize the optimal dual solution. The directional derivatives of v must now be expressed in terms of the subgradients of f and g. This can be approached by considering how the problem (11.2) can be solved.

Suppose (11.2) were to be solved by a subgradient approximation algorithm similar to the generalized Benders' algorithm discussed in previous chapters. The objective function $f(\underline{x},\underline{w})$ can be represented by its subgradients. Let $\hat{\underline{\xi}} = [\hat{\underline{\xi}}_1, \hat{\underline{\xi}}_2]$ be a subgradient of f at $(\hat{\underline{x}},\hat{\underline{w}})$. Then for any $(\underline{x},\underline{w})$

$$f(\underline{x},\underline{w}) \geq f(\hat{\underline{x}},\hat{\underline{w}}) + \underline{\xi}_1(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\xi}}_2(\underline{w}-\hat{\underline{w}}).$$

Furthermore, since this relationship holds with equality when $(\underline{x},\underline{w}) = (\hat{\underline{x}},\hat{\underline{w}})$,

$$f(\underline{x},\underline{w}) = \max_{(\hat{\underline{x}},\hat{\underline{w}})} f(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\xi}}_1(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\xi}}_2(\underline{w}-\hat{\underline{w}}) \tag{11.6}$$

Similarly, each constraint function g_k can be represented by its subgradients. Let $\hat{\underline{\eta}}_k = [\hat{\underline{\eta}}_{1k}, \hat{\underline{\eta}}_{2k}]$ be a subgradient of g_k at $(\hat{\underline{x}},\hat{\underline{w}})$. Then for any $(\underline{x},\underline{w})$

$$g_k(\underline{x},\underline{w}) \geq g_k(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\eta}}_{1k}(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\eta}}_{2k}(\underline{w}-\hat{\underline{w}})$$

and $$g_k(\underline{x},\underline{w}) = \max_{(\hat{\underline{x}},\hat{\underline{w}})} g_k(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\eta}}_{1k}(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\eta}}_{2k}(\underline{w}-\hat{\underline{w}}).$$

Then $g_k(\underline{x},\underline{w}) \leq 0$ if and only if

$$0 \geq \max_{(\hat{\underline{x}},\hat{\underline{w}})} g_k(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\eta}}_{1k}(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\eta}}_{2k}(\underline{w}-\hat{\underline{w}})$$

or equivalently

$$0 \geq g_k(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\eta}}_{1k}(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\eta}}_{2k}(\underline{w}-\hat{\underline{w}}) \quad \text{for all} \quad (\hat{\underline{x}},\hat{\underline{w}}).$$

By collecting all of the component subgradients into a matrix $\underline{\eta} = [\underline{\eta}_1, \underline{\eta}_2]$, this condition can be written

$$g(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\eta}}_1(\underline{x}-\hat{\underline{x}}) + \underline{\eta}_2(\underline{w}-\hat{\underline{w}}) \leq 0 \quad \text{for all} \quad (\hat{\underline{x}},\hat{\underline{w}}) \tag{11.7}$$

Then problem (11.2) can be written in the form

$$v(\underline{y}) = \underset{z,\underline{x},\underline{w}}{\text{minimum}}\ z \tag{11.8}$$

$$\text{subject to } z \geq f(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\underline{\xi}}_1(\underline{x}-\hat{\underline{x}}) + \hat{\underline{\xi}}_2(\underline{w}-\hat{\underline{w}}) \text{ for all } (\hat{\underline{x}},\hat{\underline{w}}) \tag{11.9}$$

$$0 \geq g(\hat{\underline{x}},\hat{\underline{w}}) + \hat{\eta}_1(\underline{x}-\hat{\underline{x}}) + \hat{\eta}_2(\underline{w}-\hat{\underline{w}}) \text{ for all } (\hat{\underline{x}},\hat{\underline{w}}) \tag{11.10}$$

$$\underline{w} = \underline{y} \tag{11.11}$$

In practice this problem would be solved by the relaxation strategy used to solve the generalized Benders' master problem. Given a set of previous trial solutions $(\underline{x}^i,\underline{w}^i)$ and their associated subgradients $(\underline{\xi}^i,\underline{\eta}^i)$, with $i = 1,\ldots,n$, a relaxed version of the problem is solved for a new trial solution. The relaxed problem is the linear program

$$\underset{z,\underline{x},\underline{w}}{\text{minimize}}\ z \tag{11.12}$$

$$\text{subject to } z \geq f(\underline{x}^i,\underline{w}^i) + \underline{\xi}_1^i(\underline{x}-\underline{x}^i) + \underline{\xi}_2^i(\underline{w}-\underline{w}^i) \quad i=1,\ldots,n \tag{11.13}$$

$$0 \geq g(\underline{x}^i,\underline{w}^i) + \underline{\eta}_1^i(\underline{x}-\underline{x}^i) + \underline{\eta}_2^i(\underline{w}-\underline{w}^i) \quad i=1,\ldots,n \tag{11.14}$$

$$\underline{w} = \underline{y} \tag{11.15}$$

Since the relaxed problem omits many of the constraints of the original problem, the new trial solution may not be feasible in the original problem (11.8) - (11.11). Thus it is necessary to determine for the new trial solution if any constraint of form (11.9) or (11.10) is violated, and if so, new constraints must be included in the relaxed problem. In order to determine which of the constraints not already included in the relaxed problem are violated by the current trial solution, a subproblem would be solved and this subproblem would also generate the required subgradients. This subproblem is analogous to the operating subproblem of the capacity planning model, derived previously. The details do not need to be disc··sed here.

Suppose that after a finite number of trial solutions have been generated, one is found which is optimal in the original problem. (In general, finite convergence cannot be guaranteed; however, finite convergence can be proven when f and g are piecewise linear, as they are in the capacity planning model to be discussed in the next section. It also seems likely that the argument can be generalized to the case where convergence to within some tolerance ε in a finite number of steps can be proven.) Let n be the number of iterations required and let (11.12) - (11.15) represent the relaxed problem which yields the optimal solution. Let θ^i be the dual multiplier on the i^{th} constraint (11.13), $\underline{\nu}^i$ the multiplier on (11.14) and $\underline{\gamma}$ the multiplier on (11.15). Since (11.12) - (11.15) is a linear program it is equivalent to the dual problem

$$v(\underline{y}) = \underset{\underline{\theta},\underline{\nu},\underline{\gamma}}{\text{maximum}} \left\{ \sum_{i=1}^{n} \theta^i [f(\underline{x}^i,\underline{w}^i) - \underline{\xi}_1^i \underline{x}^i - \underline{\xi}_2^i \underline{w}^i] + \right.$$

$$\left. \sum_{i=1}^{n} \underline{\nu}^i [g(\underline{x}^i,\underline{w}^i) - \underline{\eta}_1^i \underline{x}^i - \underline{\eta}_2^i \underline{w}^i] - \underline{\gamma}\,\underline{y} \right\} \qquad (11.16)$$

subject to

$$\sum_{i=1}^{n} \theta^i = 1 \qquad (11.17)$$

$$\sum_{i=1}^{n} (\theta^i \underline{\xi}_1^i + \underline{\nu}^i \underline{\eta}_1^i) = 0 \qquad (11.18)$$

$$\sum_{i=1}^{n} (\theta^i \underline{\xi}_2^i + \underline{\nu}^i \eta_2^i) - \underline{\gamma} = 0 \qquad (11.19)$$

θ^i, $\underline{\nu}^i \geq 0$, $\underline{\gamma}$ unrestricted in sign.

By the discussion given above, $\underline{\gamma}$ is a subgradient of $v(\underline{y})$ if and only if it is an optimal multiplier vector for (11.12) - (11.15), or in other words, if it is part of an optimal solution to (11.16) - (11.19). However, by (11.19)

$$\underline{\gamma} = \sum_{i=1}^{n} (\theta^i \underline{\xi}_2^i + \underline{\nu}^i \underline{\eta}_2^i)$$

where θ^i and $\underline{\nu}^i$ are the shadow prices on constraints (11.13) and (11.14) respectively.

This procedure for finding $\underline{\gamma}$ simplifies considerably. The problem (11.12) - (11.15) is equivalent to

$$\underset{z,\underline{x}}{\text{minimize}}\ \underline{z} \tag{11.20}$$

$$\text{subject to } z \geq f(\underline{x}^i,\underline{y}) + \underline{\xi}_1^i(\underline{x}-\underline{x}^i) \quad i=1,\ldots,n \tag{11.21}$$

$$0 \geq g(\underline{x}^i,\underline{y}) + \underline{\eta}_1^i(\underline{x}-\underline{x}^i) \tag{11.22}$$

As constraints are generated in this problem the sub-gradients $\underline{\xi}_2^i$ and $\underline{\eta}_2^i$, for the second arguments of f and g, are also calculated. When the trial solution generated by this relaxed problem finally yields an optimal solution to the master problem, the optimal multipliers θ^i, on (11.21), and $\underline{\nu}^i$, on (11.22), are used to calculate

$$\underline{\gamma} = \sum_{i=1}^{n} (\theta^i\underline{\xi}_2^i + \underline{\nu}^i\underline{\eta}_2^i)$$

It should be noted that this subgradient is not unique, and that in order to compute directional derivatives, the subdifferential must be computed for use in (11.4). The method proposed here can be extended to compute the subdifferential, but it will not be pursued here.

C. Application to the Capacity Planning Problem

Recall from Chapter 10 that the capacity planning problem to be solved is

$$TC(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\underline{\varepsilon}_T) = \underset{\underline{X},\underline{Y}_1,\ldots,\underline{Y}_T}{\text{minimum}} \ \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t,\underline{Q}_t) \qquad (11.23)$$

$$\text{subject to} \quad EG_t(\underline{Y}_t,\underline{Q}_t) \le \varepsilon_t \qquad t = 1,\ldots,T \qquad (11.24)$$

$$0 \le \underline{Y}_t \le \delta_t\underline{X} \qquad (11.25)$$

and it is desired to find a subgradient of TC. In order to apply the discussion of the last section, it is necessary to associate the generalized Benders' master problem for this model with the subgradient approximation problem (11.12) - (11.15) of the previous section. It is also necessary to identify the subgradients of the functions EF_t and EG_t. When these correspondences have been made, then the required subgradient can be calculated from the shadow prices associated with the optimal master program, rather than from the nonlinear program (11.23) - (11.25) itself.

The correspondence between the generalized Benders' master problem and the subgradient approximation problem is made by showing that solving the operating subproblems generates subgradients of the objective function (11.23) and of the unserved energy function in (11.24).

Define the operating subproblem (10.10) - (10.12) in the following parametric fashion

$$EF^*(\underline{X},\underline{Q},\varepsilon) = \underset{Y^1,\ldots,Y^I}{\text{minimum}} \sum_{i=1}^{I} \sum_{s=1}^{S} \int_0^{L_s} [g_i(Q)-g_{i+1}(Q)]dQ \qquad (11.25)$$

$$\text{subject to} \quad \sum_{s=1}^{S} \int_0^{L_s} g_{I+1}(Q)dQ \leq \varepsilon \qquad (11.26)$$

$$0 \leq Y^i \leq X^i \quad i = 1,\ldots,I \qquad (11.27)$$

$$L_s = Q_s \quad s = 1,\ldots,S \qquad (11.28)$$

By the first proposition of the previous section, if optimal dual multipliers can be found which satisfy the global optimality conditions, they give a subgradient of the function EF*. The strategy here, as in Chapter 8, will be to propose optimal primal and dual solutions for the problem, and then to show thay they satisfy the global optimality conditions.

The optimal primal solution is the same as was used before

$$\hat{Y}^i = \begin{cases} X^i & i < n \\ 0 & i > n \end{cases}$$

where the marginal plant index n is defined so that

$$\sum_{s=1}^{S} \int_0^{Q_s} g_{n+1}(Q)\,dQ = \varepsilon$$

with $0 < Y^n \leq X^n$. The dual multiplier π the unserved energy constraint (11.26) is equal to the cost of operating the marginal plant

$$\hat{\pi} = F^n.$$

The shadow prices $\hat{\lambda}^i$ on the capacity constraints (11.27) are defined by the formula (9.20) given in Chapter 9,

$$\hat{\lambda}^j = \begin{cases} -\Delta_j G_j(U^j) - \sum_{i=j+1}^{n} \Delta_i p_i H_{ij}(U^i) + \Delta_n H_{n+1,j}(U^n) & \text{for } j < n \\ 0 & \text{for } j \geq n \end{cases}$$

(Since the functions in this problem are not differentiable, the Kuhn-Tucker conditions cannot be used to find λ^j. However, the validity of the formula given above will be established below). Finally, the shadow prices the identity constraints (11.28)

$$\hat{\gamma}_s = \sum_{i=1}^{n} (F^i - F^{i-1}) g_i(Q_s) \qquad s = 1,\ldots,S \tag{11.29}$$

(where $F^0 = 0$).

It is necessary to show that the proposed optimal solution $Y^i = \hat{Y}^i$ and $L_s = Q_s$ minimizes the Lagrangian of the problem

$$L(\underline{Y},\underline{L},\underline{\lambda},\underline{\gamma},\pi) =$$

$$\sum_{i=1}^{I} \sum_{s=1}^{S} F^i \int_0^{L_s} [g_i(Q) - g_{i+1}(Q)]dQ + \pi \sum_{s=1}^{S} \int_0^{L_s} g_{I+1}(Q)dQ$$

$$+ \sum_{i=1}^{I} \lambda^i (Y^i - X^i) + \sum_{s=1}^{S} \gamma_s (Q_s - L_s) \tag{11.30}$$

when the optimal dual multipliers are used $\pi = \hat{\pi}$, $\underline{\lambda} = \hat{\underline{\lambda}}$, $\underline{\gamma} = \hat{\underline{\gamma}}$. This is the same problem which was faced in Section B of Chapter 8; however, the proof given there used the differentiability of the Lagrangian to define the multipliers, which cannot be done here. The argument given below is therefore based on a subgradient property of this Lagrangian, given in the proposition below.

Define $w_i(Q^*;Y^1,\ldots,Y^{i-1}) = \int_0^{Q^*} g_i(Q;Y^1,\ldots,Y^{i-1})dQ$ and define $H_{ij}(U,Q^*)$ by the recursive formulas (9.10) given in Chapter 9.

$$H_{i+1,j}(U,Q^*) = p_i H_{ij}(U,Q^*) + q_i H_{ij}(U-Y^i,Q^*) \quad i = j+1,\dots,I$$

$$H_{j+1,j}(U,Q^*) = p_j G_j(U,Q^*)$$

where $G_j(U,Q^*)$ is the equivalent load duration curve derived from the step function load duration curve

$$G(Q,Q^*) = \begin{cases} 1 & \text{if } Q < Q^* \\ 0 & \text{if } Q \geq Q^* \end{cases}$$

<u>Second Proposition</u>: For any values Q^* and $Y^1,\dots,Y^{i-1}$

$$w_i(Q^*;Y^1,\dots,Y^{i-1}) \geq w_i(\hat{Q};\hat{Y}^i,\dots,\hat{Y}^{i-1})$$
$$+ g_i(\hat{Q})(Q^*-\hat{Q}) - \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1},\hat{Q})(Y^j-\hat{Y}^j), \qquad (11.31)$$

where $\hat{U}^{i-1} = \hat{Y}^1 + \dots + \hat{Y}^{i-1}$ and g_i and H_{ij} are defined using $\hat{Y}^1,\dots,\hat{Y}^{i-1}$.

<u>Proof</u>: The proof is by induction on i. When $i = 1$, $w_1(\hat{Q}) = \max(0,\hat{Q})$. It is to be shown that

$$w_1(Q^*) \geq w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q})$$

There are four possible cases:

i) If $Q^* > 0$ and $\hat{Q} > 0$, then $w_1(Q^*) = Q^*$ and

$w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q}) = \hat{Q} + 1\cdot(Q^*-\hat{Q}) = Q^*$. Hence
$w_1(Q^*) = w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q})$.

ii) If $Q^* \leq 0$ and $\hat{Q} \leq 0$, then $w_1(Q^*) = 0$ and $w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q}) = 0 + 0\cdot(Q^*-\hat{Q}) = 0$. Hence equality holds once again in (11.31).

iii) If $Q^* > 0$ and $\hat{Q} \leq 0$, then $w_1(Q^*) = Q^*$ and $w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q}) = 0 + 0\cdot(Q^*-\hat{Q}) = 0$. Since $Q^* > 0$, the inequality (11.31) holds.

iv) If $Q^* \leq 0$ and $\hat{Q} > 0$, then $w_1(Q^*) = 0$ and $w_1(\hat{Q}) + g_1(\hat{Q})(Q^*-\hat{Q}) = \hat{Q} + 1\cdot(Q^*-\hat{Q}) = Q^*$. Since $0 \geq Q^*$, the inequality (11.31) holds.

Now suppose the inequality holds for some i. It must be shown that it holds also for i+1.

$$
\begin{aligned}
w_{i+1}(Q^*;Y^1,\ldots,Y^i) = \\
&p_i w_i(Q^*-Y^i;Y^1,\ldots,Y^{i-1}) \\
&+ q_i w_i(Q^*;Y^1,\ldots,Y^{i-1}) \\
\geq p_i\{&w_i(\hat{Q}-\hat{Y}^i;\hat{Y}^1,\ldots,\hat{Y}^{i-1}) + g_i(\hat{Q}-\hat{Y}^i)(Q^*-\hat{Q}-Y^i+\hat{Y}^i) \\
&- \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1},\hat{Q}-\hat{Y}^i)(Y^j-\hat{Y}^j)\}
\end{aligned}
$$

$$+ q_i\{w_i(\hat{Q};\hat{Y}^1,\ldots,\hat{Y}^{i-1}) + g_i(\hat{Q})(Q^*-\hat{Q})$$

$$- \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1},\hat{Q})(Y^j-\hat{Y}^j)$$

$$= w_{i+1}(\hat{Q};\hat{Y}^1,\ldots,\hat{Y}^i) + g_{i+1}(\hat{Q})(Q^*-\hat{Q})$$

$$- p_i g_i(\hat{Q}-\hat{Y}^i)(Y^i-\hat{Y}^i) - \sum_{j=1}^{i-1} \{p_i H_{ij}(\hat{U}^{i-1},\hat{Q}-\hat{Y}^i)$$

$$+ q_i H_{ij}(\hat{U}^{i-1},\hat{Q})\}(Y^j-\hat{Y}^j)$$

Now, it was shown in Section 10.C that

$$G_i(Q,\hat{Q}+Y) = G_i(Q-Y,\hat{Q}),$$

and H_{ij} inherits this property as well. Furthermore,

$$g_1(\hat{Q}-Q) = \begin{cases} 1 & \text{if } \hat{Q} - Q > 0 \\ 0 & \text{if } \hat{Q} - Q \le 0 \end{cases}$$

$$= \begin{cases} 1 & \text{if } Q < \hat{Q} \\ 0 & \text{if } Q \ge \hat{Q} \end{cases}$$

$$= G_1(Q,\hat{Q})$$

and by induction

$$g_i(\hat{Q}-Q) = G_i(Q+\hat{U}^{i-1},\hat{Q}).$$

Therefore

$$- p_i g_i(\hat{Q}-\hat{Y}^i)(Y^i-\hat{Y}^i) - \sum_{j=1}^{i-1} \{p_i H_{ij}(\hat{U}^{i-1},\hat{Q}-\hat{Y}^i)$$

$$+ q_i H_{ij}(\hat{U}^{i-1},\hat{Q})\}(Y^j-\hat{Y}^j)$$

$$= - p_i G_i(\hat{U}^i,\hat{Q})(Y^i-\hat{Y}^i) - \sum_{j=1}^{i-1} \{p_i H_{ij}(\hat{U}^i,\hat{Q})$$

$$+ q_i H_{ij}(\hat{U}^i-\hat{Y}^i,\hat{Q})\}(Y^j-\hat{Y}^j)$$

$$= - \sum_{j=1}^{i} H_{i+1,j}(U^i,\hat{Q})(Y^j-\hat{Y}^j)$$

and therefore

$$w_{i+1}(Q^*;Y^1,\ldots,Y^i) \geq w_{i+1}(\hat{Q};\hat{Y}^1,\ldots,\hat{Y}^i)$$

$$+ g_{i+1}(\hat{Q})(Q^*-\hat{Q}) - \sum_{j=1}^{i} H_{i+1,j}(\hat{U}^i,\hat{Q})(Y^j-\hat{Y}^j). \;||$$

This proposition gives a subgradient relationship for the function

$$\int_0^{\hat{Q}} g_i(Q;\hat{Y}^1,\ldots,\hat{Y}^{i-1})dQ$$

The next proposition, similar to one proved in Section B of Chapter 8, shows that $Y^i = \hat{Y}^i$, $\hat{L}_s = Q_s$ minimizes part of the Lagrangian.

<u>Third Proposition</u>: The proposed optimal solution minimizes the function

$$\sum_{i=1}^{n} F^i \int_0^{L_s} [g_i(Q)-g_{i+1}(Q)]dQ + \hat{\pi} \int_0^{L_s} g_{n+1}(Q)dQ$$

$$+ \sum_{i=1}^{n} \hat{\lambda}^i(Y^i-X^i) + \hat{\gamma}_s(Q_s-L_s) \qquad (11.32)$$

Proof: Since $\hat{\pi} = F^n$, this expression can be written

$$\sum_{i=1}^{n} (F^i-F^{i-1}) \int_0^{L_s} g_i(Q)dQ + \sum_{i=1}^{n} \hat{\lambda}^i(Y^i-X^i) + \hat{\gamma}_s(Q_s-L_s)$$

$$= \sum_{i=1}^{n} (F^i-F^{i-1})w_i(L_s;Y^1,\ldots,Y^{i-1}) + \sum_{i=1}^{n} \hat{\lambda}^i(Y^i-X^i) + \hat{\gamma}_s(Q_s-L_s)$$

By the preceding proposition, for arbitrary values of Y^i and L_s, and since $F^i \geq F^{i-1}$,

$$\sum_{i=1}^{n} (F^i-F^{i-1})w_i(L_s;Y^1,\ldots,Y^{i-1}) \geq$$

$$\sum_{i=1}^{n} (F^i-F^{i-1})w_i(Q_s;\hat{Y}^1,\ldots,\hat{Y}^{i-1}) + \sum_{i=1}^{n} (F^i-F^{i-1})g_i(Q_s)(L_s-Q_s)$$

$$- \sum_{i=1}^{n} (F^i-F^{i-1}) \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1},Q_s)(Y^j-\hat{Y}^j)$$

Now the final term is

$$\sum_{i=1}^{n} (F^i - F^{i-1}) \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1}, Q_s)(Y^j - \hat{Y}^j)$$

$$= \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (F^i - F^{i-1}) H_{ij}(\hat{U}^{i-1}, Q_s)(Y^j - \hat{Y}^j)$$

$$= \sum_{j=1}^{n-1} \sum_{i=j+1}^{j} (F^i - F^{i-1}) \{\beta_{i-1,j} p_j G_j(\hat{U}^j, Q_s) +$$

$$\sum_{k=j+1}^{i-1} \beta_{i-1,k} p_k H_{kj}(\hat{U}^k, Q_s)\}(Y^j - \hat{Y}^j)$$

(as was shown in Chapter 9. β_{ik} is defined in (9.16))

$$= \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (F^i - F^{i-1}) \beta_{i-1,j} p_j G_j(\hat{U}^j, Q_s)(Y^j - \hat{Y}^j)$$

$$+ \sum_{j=1}^{n-1} \sum_{k=j+1}^{n-1} \sum_{i=k+1}^{n} (F^i - F^{i-1}) \beta_{i-1,k} p_k H_{kj}(\hat{U}^k, Q_s)(Y^j - \hat{Y}^j)$$

Now

$$\sum_{i=k+1}^{n} (F^i - F^{i-1}) \beta_{i-1,k} p_k$$

$$= \sum_{i=k+1}^{n} F^i (\beta_{i-1,k} - \beta_{ik}) p_k - F^k p_k + F^n \beta_{n-1,k} p_k$$

$$= \sum_{i=k+1}^{n} F^i p_i \beta_{i-1,k} p_k - F^k p_k + F^n \beta_{n-1,k} p_k$$

$$= \sum_{i=k+1}^{I} F^i p_i \beta_{i-1,k} p_k - F^k p_k$$

$$- \sum_{i=n+1}^{I} F^i p_i \beta_{i-1,k} p_k + F^n \beta_{n-1,k} p_k$$

$$= - \Delta_k p_k + \Delta_n p_k \beta_{nk}$$

(by the formulas given in Chapter 9)

Hence

$$\sum_{i=1}^{n} (F^i - F^{i-1}) \sum_{j=1}^{i-1} H_{ij}(\hat{U}^{i-1}, Q_s)(Y^j - \hat{Y}^j)$$

$$= \sum_{j=1}^{n-1} (-\Delta_j p_j + \Delta_n p_j \beta_{nj}) G_j(\hat{U}^j, Q_s)(Y^j - \hat{Y}^j)$$

$$+ \sum_{j=1}^{n-1} \sum_{k=j+1}^{n-1} (-\Delta_k p_k + \Delta_n p_k \beta_{nk}) H_{kj}(\hat{U}^k, Q_s)(Y^j - \hat{Y}^j)$$

$$= - \sum_{j=1}^{n-1} \{\Delta_j p_j G_j(\hat{U}^j, Q_s) + \sum_{k=j+1}^{n-1} \Delta_k p_k H_{kj}(\hat{U}^k, Q_s)\}(Y^j - \hat{Y}^j)$$

$$+ \Delta_n \sum_{j=1}^{n-1} \{p_j \beta_{nj} G_j(\hat{U}^j, Q_s) + \sum_{k=j+1}^{n-1} p_k \beta_{nk} H_{kj}(\hat{U}^k, Q_s)\}(Y^j - \hat{Y}^j)$$

$$= - \sum_{j=1}^{n-1} \{\Delta_j p_j G_j(\hat{U}^j, Q_s) + \sum_{k=j+1}^{n} \Delta_k p_k H_{kj}(\hat{U}^k, Q_s)\}(Y^j - \hat{Y}^j)$$

$$+ \Delta_n \sum_{j=1}^{n-1} H_{n+1,j}(\hat{U}^j, Q_s)(Y^j - \hat{Y}^j)$$

$$= \sum_{j=1}^{n-1} \hat{\lambda}^j (Y^j - \hat{Y}^j)$$

Hence, finally, since $(F^i - F^{i-1}) g_i(Q_s) = \hat{\gamma}_s$

$$\sum_{i=1}^{n} (F^i - F^{i-1}) w_i(L_s; Y^1, \ldots, Y^{i-1}) \geq$$

$$\sum_{i=1}^{n} (F^i - F^{i-1}) w_i(Q_s; \hat{Y}^1, \ldots, \hat{Y}^{i-1}) + \hat{\gamma}_s(L_s - Q_s)$$

$$- \sum_{i=1}^{n} \hat{\lambda}^i (Y^i - \hat{Y}^i)$$

Now, therefore, for any L_s, $Y^1, \ldots, Y^n$,

$$\sum_{i=1}^{n} F^i \int_0^{L_s} [g_i(Q) - g_{i+1}(Q)]dQ + \hat{\pi} \int_0^{L_s} g_{n+1}(Q)dQ$$

$$+ \sum_{i=1}^{n} \hat{\lambda}^i (Y^i - X^i) + \hat{\gamma}_s(Q_s - L_s)$$

$$\geq \sum_{i=1}^{n} F^i \int_0^{Q_s} [g_i(Q) - g_{i+1}(Q)]dQ + \hat{\pi} \int_0^{Q_s} g_{n+1}(Q)dQ$$

$$- \sum_{i=1}^{n} \hat{\lambda}^i (Y^i - \hat{Y}^i) + \sum_{i=1}^{n} \hat{\lambda}^i (Y^i - X^i) + \hat{\gamma}_s(L_s - Q_s)$$

$$+ \hat{\gamma}_s(Q_s - L_s)$$

$$= \sum_{i=1}^{n} F^i \int_0^{Q_s} [g_i(Q) - g_{i+1}(Q)]dQ + \hat{\pi} \int_0^{Q_s} g_{n+1}(Q)dQ$$

$$+ \sum_{i=1}^{n} \hat{\lambda}^i (\hat{Y}^i - X^i)$$

which is the value of the function (11.32) evaluated with $Y^i = \hat{Y}^i$ and $L_s = Q_s$. Hence these values minimize the function. ||

An argument similar to the one given in the second proposition of Section 8.D shows that $\hat{Y}^i = 0$ for $i > n$ must be true at a local minimum of the Lagrangian. Since the Lagrangian then reduces to the function (11.32) given above, the proposed optimal solution is the minimal solution and thus the proposed primal and dual solutions satisfy the global optimality conditions. Then, by the first proposition of the previous section, $\hat{\underline{\lambda}}$, $\hat{\underline{\gamma}}$, $\hat{\pi}$ form a subgradient of $EF^*(\underline{X},\underline{\Omega},\varepsilon)$ with respect to $\underline{X}$, $\underline{\Omega}$, and ε.

If the subproblem (11.25) - (11.28) is infeasible for the given values of $\underline{X}$, $\underline{\Omega}$, ε, then the solution generated is

$$\hat{Y}^i = X^i \quad \text{for} \quad i = 1,\ldots,I$$

$$\hat{\pi} = F^I$$

$$\hat{\mu}^i = H_{I+1,i}(\hat{U}^I)$$

and the shadow price on the identity constraint is

$$\hat{\zeta}_s = g_{I+1}(\Omega_s).$$

The second proposition, above, also shows that for any L_s, $Y^1,\ldots,Y^I$

$$\int_0^{L_s} g_{I+1}(Q;Y^1,\ldots,Y^I)dQ \geq$$

$$\int_0^{Q_s} g_{I+1}(Q;\hat{Y}^1,\ldots,\hat{Y}^I)dQ + g_{I+1}(Q_s)(L_s-Q_s)$$

$$- \sum_{j=1}^{I} H_{I+1,j}(\hat{U}^I,Q_s)(Y^j-\hat{Y}^j)$$

and thus, in this case also, the subproblem multipliers give a subgradient.

Now, the generalized Benders' master problem for (11.23) - (11.25) is

$$\underset{\underline{X},\underline{Y}_1,\ldots,\underline{Y}_T}{\text{minimize}} \quad Z \tag{11.33}$$

$$\text{subject to } Z \geq \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t^k,\underline{\Omega}_t) + \underline{\lambda}_t^k\delta_t(\underline{X}^k-\underline{X}) \tag{11.34}$$

$$k = 1,\ldots,K$$

$$\sum_{t\varepsilon\Gamma^k} [EG_t(\underline{Y}_t^k,\underline{\Omega}_t) + \underline{\mu}_t^k\delta_t(\underline{X}^k-\underline{X})-\varepsilon_t] \leq 0 \tag{11.35}$$

$$X \geq 0.$$

If the subgradients $\underline{\xi}_1^k$ and $\underline{\eta}_1^k$ of the problem (11.20) - (11.22) in the previous section are associated as

$$\underline{\xi}_1^k = \sum_{t=1}^{T} \underline{\lambda}_t^k \delta_t$$

$$\underline{\eta}_1^k = \sum_{t\epsilon\Gamma^k} \underline{\mu}_t^k \delta_t$$

and if the variables $\underline{x}$ are associated with $\underline{X}$ and $\underline{y}$ (or $\underline{w}$) with $\underline{Q}_1,\ldots,\underline{Q}_T$ and $\varepsilon_1,\ldots,\varepsilon_T$, then (11.33) - (11.35) takes the general form discussed in the previous section, (11.20) - (11.22). Then, in order to calculate a subgradient of TC, it is necessary to identify the subgradients $\underline{\xi}_2^k$ and $\underline{\eta}_2^k$, for the variables $\underline{Q}_t$ and ε_t. From the discussion given above, it is clear that the component of $\underline{\xi}_2^k$ associated with Q_{ts} is

$$\gamma_{ts}^k = \sum_{i=1}^{I_t} (F^{it}-F^{i-1,t}) g_i(Q_{ts};\underline{Y}_t^k)$$

and with ε_t is π_t^k (which is the operating cost of the marginal plant in period t at iteration k). The component of $\underline{\eta}_2^k$ associated with Q_{ts} is

$$\zeta_{ts}^k = g_{I+1}(Q_{ts};\underline{Y}_t^k)$$

and with ε_t is 1. (The notation $g_i(Q,\underline{Y}_t^k)$ means that g_i is calculated using the values of Y from the k^{th} iteration, in period t.)

Finally, if θ^k is the shadow price associated with the k^{th} constraint (11.34) and ν^k with the k^{th} constraint (11.35), then the result of the previous section shows that a subgradient of TC is formed as a weighted combination of these subgradients. The component of this subgradient associated with Q_{ts} is

$$\sum_{k=1}^{K} \{\theta^k [\sum_{i=1}^{I_t} (F^{it}-F^{i-1,t}) g_i(Q_{ts};\underline{Y}_t^k)] + \nu^k g_{I_t+1}(Q_{ts};\underline{Y}_t^k)\}, \quad (11.36)$$

and with ε_t is

$$\sum_{k=1}^{K} \{\theta^k \pi_k^t + \nu^k\} \quad (11.37)$$

Computationally, these formulas require that, as the subproblems are solved at each iteration k, the functions $g_i(Q;\underline{Y}_t^k)$ also be computed and saved. When the generalized Benders' master problem has converged, the shadow prices on the constraints θ^k and ν^k are used to compute the subgradient components given above. Chapter 12 will discuss the use of this subgradient in computing peak-load prices.

CHAPTER 12

CALCULATING PEAK-LOAD PRICES

A. Introduction

The previous chapters of Part Three have presented a technique for using the probabilistic capacity expansion planning model to calculate marginal costs associated with the demand for electricity at different times. However, it has been noted by several authors (see, for example, Crew and Kleindorfer[1]) that it is not enough to consider system planning and pricing separately. The demand for electricity at different times depends on the price charged, and planners must consider how price will affect demand. The capacity expansion and pricing decisions must be made jointly in the context of price-elastic demand. This chapter presents an approach to integrating the marginal cost information derived in the previous chapters into a pricing model.

The model presented in this chapter is not intended to consider all the aspects of the peak-load pricing problem which have been discussed in the literature[2], nor is it necessarily practical for implementation. Instead, it is intended to demonstrate how the capacity planning model and the marginal cost information developed in previous chapters can be used in models for determining peak-load prices. There remain barriers to practical implementation of models to determine peak-load prices, some of which will be discussed in this chapter. Nevertheless, the model presented here can serve as a guide to further research and to data gathering experiments for the design of practical peak-load pricing models.

There are two main issues involved in designing a model for calculating peak-load prices. One issue is specifying the supply and the demand models. The supply model represents the relationship between quantity supplied and cost; in the model presented here, the capacity planning model serves as the supply model. The demand model represents the relationship between quantity demanded and price. There will be some discussion of demand models in the next section; however, it will generally be assumed that a relationship between demand and price is known, and its exact structure need not be specified here.

The second issue, the one which will mainly be addressed here, is modeling how supply and demand interact to determine prices. Prices are determined by the conditions for equilibrium in the marketplace. Classically, these conditions have been stated as marginal conditions, requiring differentiability of the cost functions involved. However, it is well-known that the equilibrium conditions can be converted into an optimization problem[3]. When so stated, the differentiability conditions can be relaxed. In this chapter, the peak-load pricing problem will be formulated as an optimization, and it will be shown that subgradient information, developed in the previous chapters of this part, can be used in place of derivative information in determining the prices.

An important reason for integrating the capacity planning model discussed previously in this thesis into a peak-load pricing model is that the capacity planning model is sufficiently rich in structure to realistically represent the supply alternatives available to utilities. Much of the peak-load pricing research done previously has used comparatively simply supply models. While these simple models have served well to advance the theory of peak-load pricing, more realistic representations will be required for the actual implementation of peak load pricing schemes. The purpose of this chapter is to provide a technique for using these more complex supply models in determining peak-load prices.

In Section B, a formulation for an equilibrium model for determining peak-load prices is presented. The formulation is based on the decomposition techniques discussed previously in this thesis. In this model, the capacity planning model is used as a subproblem which calculates the cost of supplying given demand for electricity. An output of this subproblem is marginal cost information, in the form of subgradients of the cost function. The master problem generates trial solutions for the equilibrium demand, using the demand model and the supply information from the subproblem.

In Section C, some of the practical difficulties of implementing the proposed peak-load pricing model are discussed. One of the major barriers is the lack of data on time-of-day price sensitivity of demand. The use of the proposed model in structuring research and experiments on peak-load pricing is discussed.

Peak-load pricing has received considerable attention in the economic research literature. Much of the traditional work[4] has used models where there is only a single generating technology or a single plant. More recent work[5] has extended the theory to consider diverse generating technologies and multiple plants. A recent paper by Crew and Kleindorfer[6] comes the closest to the work presented here. They consider diverse technologies and uncertain demand (rather than supply) and derive equilibrium conditions for peak-load prices and capacity expansion. However, these models have, so far, not considered models of supply with the richness of the capacity planning models presented here.

B. An Equilibrium Model for Peak-Load Pricing

Since electric utilities are regulated monopolies, models of their pricing decisions can be based on either of two objectives - profit maximization, which leads to "monopolistic" pricing decisions, or welfare maximization, which leads to "competitive" pricing decisions. In using profit maximization, the utility is generally subject to a constraint on allowed rate of return on invested capital, which serves to move the prices away from the unconstrained monopoly solution toward the competitive solution. On the other hand, in using welfare maximization, the model attempts to simulate the competitive market situation to determine what the competitive solution would be if there were, in fact, a competitive market for electricity. The model presented in this chapter will use welfare maximization as the objective, following much of the recent literature on peak-load pricing. This criterion seems to be more acceptable to the regulatory authorities, as representatives of the public at large, and avoids some of the supposed biases of rate-of-return-constrained profit maximization. The difficulty with it is that it requires information about demand which may be difficult or impossible to observe in practice.

The welfare maximization approach assumes the existence of a consumers' surplus function[7], $CS(\underline{Q})$, which measures the benefit to consumers associated with a demand of $\underline{Q}$. Given a cost function $TC(\underline{Q})$, which measures the cost to producers of supplying $\underline{Q}$, the problem of welfare maximization is expressed as

$$\underset{\underline{Q}}{\text{maximize}}\ CS(\underline{Q}) - TC(\underline{Q}). \tag{12.1}$$

Assuming, for the moment, that both CS and TC are differentiable and $\underline{Q}$ is unconstrained, the necessary condition for optimality in this problem is that

$$\frac{\partial CS}{\partial Q_i} = \frac{\partial TC}{\partial Q_i} \quad \text{for all } i.$$

The derivative $\frac{\partial TC}{\partial Q_i}$ is just the marginal cost function for commodity i. The derivative $\frac{\partial CS}{\partial Q_i}$ is the inverse of the demand function for commodity i, which gives the price consumers are willing to pay for commodity i. Thus, this optimality condition is just the usual equilibrium condition for a competitive market, that price (as given by the market demand curve) is equal to marginal cost. This formulation will now be generalized to the case when the cost function TC is given by an optimization problem, such as has been discussed in previous chapters. It will be assumed that the consumers' surplus function is available (the difficulties in measuring this function will be discussed in the next section).

As before, let $\underline{Q}_t$ represent a vector of hourly demands Q_{ts} during year t, and let ε_t be the reliability of service (measured by expected unserved energy) in year t. Let $CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T)$ be the consumers' surplus associated with the given demands and reliabilities (often, $CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) = CS(\underline{Q}_1,\ldots,\underline{Q}_T) - R(\varepsilon_1,\ldots,\varepsilon_T)$, where R is a <u>rationing</u> cost function, as discussed in Section D of Chapter 4). As shown in Chapter 10, the cost of meeting the given demands and reliabilities is given by the capacity planning model (10.13) - (10.15)

$$TC(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) = \underset{\underline{X},\underline{Y}_1,\ldots,\underline{Y}_T}{\text{minimum}} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t,\underline{Q}_t)$$

$$\text{subject to} \quad EG_t(\underline{Y}_t,\underline{Q}_t) \le \varepsilon_t \qquad t=1,\ldots,T \tag{12.2}$$

$$0 \le \underline{Y}_t \le \delta_t \underline{X}$$

The welfare maximization problem can then be written

$$\underset{\substack{\underline{Q}_1,\ldots,\underline{Q}_T \\ \varepsilon_1,\ldots,\varepsilon_T \\ \underline{Y}_1,\ldots,\underline{Y}_T \\ \underline{X}}}{\text{maximize}} \quad CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) - \underline{C}'\underline{X} - \sum_{t=1}^{T} EF_t(\underline{Y}_t,\underline{Q}_t)$$

$$\begin{aligned} \text{subject to} \quad & EG_t(\underline{Y}_t,\underline{Q}_t) \le \varepsilon_t \\ & \qquad\qquad t = 1,\ldots,T \\ & 0 \le \underline{Y}_t \le \delta_t\underline{X} \\ & \underline{Q}_t \ge 0,\ \varepsilon_t \ge 0 \end{aligned} \tag{12.3}$$

As with the models discussed previously in this thesis, this model can be solved by a decomposition approach[8]

$$\underset{\substack{\underline{Q}_1,\ldots,\underline{Q}_T \ge 0 \\ \varepsilon_1,\ldots,\varepsilon_T \ge 0}}{\text{maximize}} \left\{ CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) - \left| \begin{aligned} & \underset{\underline{X},\underline{Y}_1,\ldots,\underline{Y}_T}{\text{minimum}} \quad \underline{C}'\underline{X} + \sum_{t=1}^{T} EF_t(\underline{Y}_t,\underline{Q}_t) \\ & \text{subject to } EG_t(\underline{Y}_t,\underline{Q}_t) \le \varepsilon_t \\ & \qquad\qquad 0 \le \underline{Y}_t \le \delta_t\underline{X} \end{aligned} \right| \right\} \tag{12.4}$$

where the inner minimization just gives the function TC. As has been suggested before, the function TC can be approximated by its subgradients. That is the problem above can be written

$$\underset{\substack{\underline{Q}_1,\ldots,\underline{Q}_T \geq 0 \\ \varepsilon_1,\ldots,\varepsilon_T \geq 0 \\ Z}}{\text{maximize}} \quad CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) - Z \tag{12.5}$$

$$\text{subject to} \quad Z \geq TC(\hat{\underline{Q}}_1,\hat{\varepsilon}_1,\ldots,\hat{\underline{Q}}_T,\hat{\varepsilon}_T)$$

$$+ \sum_{t=1}^{T} [\hat{\underline{\gamma}}_t(\underline{Q}_t-\hat{\underline{Q}}_t) + \hat{\pi}_t(\varepsilon_t-\hat{\varepsilon}_t)] \tag{12.6}$$

$$\text{for all } (\hat{\underline{Q}}_1,\hat{\varepsilon}_1,\ldots,\hat{\underline{Q}}_T,\hat{\varepsilon}_T)$$

where $\hat{\underline{\gamma}}_t$ is a subgradient of TC with respect to $\underline{Q}_t$ at $\hat{\underline{Q}}_t$ and $\hat{\pi}_t$ is a subgradient with respect to ε_t at $\hat{\varepsilon}_t$. Both the values of TC and of the subgradients $\underline{\gamma}$ and π at a specific trial value $(\hat{Q}_1,\varepsilon_1,\ldots,\hat{\underline{Q}}_T,\varepsilon_T)$ can be found by solving the capacity planning problem (12.2) as described in Chapters 10 and 11. In actual computation, this problem could be solved by a relaxation strategy in which the constraints (12.6) are generated successively. A relaxed master problem would be solved to generate a trial solution. This relaxed problem has the form

$$\underset{\substack{\underline{Q}_1,\ldots,\underline{Q}_T \geq 0 \\ \varepsilon_1,\ldots,\varepsilon_T \geq 0 \\ Z}}{\text{maximize}} \quad CS(\underline{Q}_1,\varepsilon_1,\ldots,\underline{Q}_T,\varepsilon_T) - Z \tag{12.7}$$

$$\text{subject to} \quad Z \geq TC(\underline{Q}_1^k,\varepsilon_1^k,\ldots,\underline{Q}_t^k,\varepsilon_T^k) +$$

$$+ \sum_{t=1}^{T} [\underline{\gamma}_t^k(\underline{Q}_t-\underline{Q}_t^k) + \pi_t^k(\varepsilon_t-\varepsilon_t^k)] \quad k=1,\ldots,K \qquad (12.8)$$

where k indexes the trial soltuion generated at iteration k. Using the $K+1^{\underline{st}}$ trial solution generated by solving this problem, the capacity planning problem (12.2) is solved to evaluate TC and to determine the subgradients, to form a new constraint in the relaxed master problem. This procedure continues, alternately solving the master and subproblems, until an optimal, or near-optimal, solution has been found.

Notice that the proposed optimization model does not determine prices directly. Instead, it determines the optimal quantities of electricity to be supplied, $\underline{Q}_t$, and the optimal reliability levels, ε_t. The prices can be determined from the optimality conditions on the master problem. Suppose that an optimal solution is found at iteration K. Let θ^k be the dual multiplier associated with the $k^{\underline{th}}$ constraint (12.8). Then the Kuhn-Tucker conditions for the problem (12.7) - (12.8) give

i) $\dfrac{\partial CS}{\partial Q_{ts}} - \sum_{k=1}^{K} \theta^k \gamma_{ts}^k \geq 0$ with equality

when $Q_{ts} > 0$

AD-A058 421 MASSACHUSETTS INST OF TECH CAMBRIDGE OPERATIONS RESE--ETC F/G 12/2
DECOMPOSITION AND PROBABILISTIC SIMULATION IN ELECTRIC UTILITY --ETC(U)
AUG 78 J A BLOOM DAAG29-76-C-0064
UNCLASSIFIED TR-154 ARO-14261.10-M NL

4 OF 4
AD
A058421

END
DATE
FILMED
10-78
DDC

OF 4

8421

$$+ \sum_{t=1}^{T} [\underline{\gamma}_t^k(\underline{Q}_t - \underline{Q}_t^k) + \pi_t^k(\varepsilon_t - \varepsilon_t^k)] \quad k = 1,\ldots,K \quad (12.8)$$

where k indexes the trial soltuion generated at iteration k. Using the K+1$\underline{\text{st}}$ trial solution generated by solving this problem, the capacity planning problem (12.2) is solved to evaluate TC and to determine the subgradients, to form a new constraint in the relaxed master problem. This procedure continues, alternately solving the master and subproblems, until an optimal, or near-optimal, solution has been found.

Notice that the proposed optimization model does not determine prices directly. Instead, it determines the optimal quantities of electricity to be supplied, $\underline{Q}_t$, and the optimal reliability levels, ε_t. The prices can be determined from the optimality conditions on the master problem. Suppose that an optimal solution is found at iteration K. Let θ^k be the dual multiplier associated with the k$\underline{\text{th}}$ constraint (12.8). Then the Kuhn-Tucker conditions for the problem (12.7) - (12.8) give

i) $\dfrac{\partial CS}{\partial Q_{ts}} - \sum_{k=1}^{K} \theta^k \gamma_{ts}^k \geq 0$ with equality

when $Q_{ts} > 0$

ii) $\frac{\partial CS}{\partial \varepsilon_t} - \sum_{k=1}^{K} \theta^k \pi_t^k \geq 0$ with equality when $\varepsilon_t > 0$

(generally, ε_t must be strictly positive for feasibility in the capacity planning model - it is impossible to reduce the unserved energy to zero at finite cost unless there is a perfectly reliable plant)

iii) $\sum_{k=1}^{K} \theta^k = 1, \quad \theta^k \geq 0$

In condition (i), the derivative $\frac{\partial CS}{\partial Q_{ts}}$ gives the price of electricity in hour s of period t, by the definition of the consumers' surplus function. Thus, this condition can be used to compute the prices as a weighted combination of the subgradients generated by various trial solutions. This condition is the analogue, for this model, of the usual "price = marginal cost" condition. Condition (ii) can be interpreted in a similar manner. It specifies that the marginal benefit of increased reliability (or the marginal reduction in rationing costs) must equal the marginal cost of providing increased reliability.

The relaxation strategy is not the only algorithm for solving this problem. A number of alternative algorithms for equilibrium problems of this type have been discussed by Shapiro[9]. They are a steepest ascent algorithm based on subgradients and a primal-dual algorithm. Both algorithms make use of the subgradient information in a somewhat more sophisticated way by using them to determine directions of change for the variables which improve the objective function. However, the specific algorithm proposed to solve the problem is less important here than the idea that the subgradient information is of key importance in solving the problem. The subgradients can be used even though marginal costs may not be uniquely defined.

C. Some Practical Issues

This section considers some of the practical issues involved in actually implementing the peak-load pricing model described in the previous section. The key issues mostly concern the specification of a demand model, which is embodied in the consumers' surplus function, and in gathering data sufficient to estimate the parameters of this model. This section will not address the practicality or the advantages and disadvantages of peak-load pricing schemes themselves, but only the issues involved in designing a model for calculating peak-load prices.

The practical issues of calculating peak-load prices have recently begun to receive a great deal of attention, as regulatory bodies in many states move toward implementation of peak-load tariffs. A major study[10] is now underway by the Electric Power Research Institute (EPRI) for the National Association of Regulatory Utility Commissioners of

> "the technology and cost of time-of-day metering and electronic methods of controlling peak-period usage of electricity and also ... of the feasibility and cost of shifting various types of usage from peak to off-peak periods."

This study includes a detailed look at time-differentiated pricing schemes, which has included study of practical methods for calculating peak-load prices.

The Rate Design Study suggests some of the major issues in the design of peak-load rate schedules. These include

i) The selection of rating periods

ii) The calculation of marginal costs

iii) The allocation of marginal costs to rating periods

iv) The elasticity of demand

Of these four issues, methods for dealing with the second and third have been discussed in depth in the earlier chapters of this part. The remaining two will be discussed here.

The model discussed previously in this part has assumed that prices will be set individually for each hour of the year; however, it is clearly impractical to do so. The data required for such a model could not be practically collected, and the number of variables in the models would be enormous. Furthermore, customers could not possibly deal with so many different prices. In addition, since many hours have similar patterns of demand

and cost, the prices would not vary considerably within groups with similar characteristics. It is therefore necessary to define a reasonably small number of types of pricing periods, with the same price prevailing during all hours which are included in a given type. Typically, four types of pricing periods might be used - say, summer peak, summer off-peak, winter peak, and winter off-peak.

Using a small number of rating periods reduces both the amount of data required and the size of the model. Furthermore, for statistical purposes, it might be assumed that observations of demand in all hours of a given type of rating period are independent observations of the same random variable, thus effectively reducing the number of variables to be observed for estimation purposes. However, reducing the number of prices to be considered induces some loss of welfare because it requires departures from the marginal cost prices which would prevail if prices were set individually for each hour (this loss is probably very slight).

The major barrier to implementing a practical peak-load pricing model is the lack of data on the price elasticity of demand for electricity at different times of the day. As has been observed in a recent survey by Taylor [12], there have been many econometric studies of demand for electricity, but, by and large, these studies have dealt with estimating annual demand for electric energy, usually by customer class (residential, commercial, or industrial). Because the price of electricity has never been charged on the basis of the time of day at which it is consumed, there is almost no data available on the price elasticity of demand by time of day. Studies are underway, however, sponsored by the Department of Energy, to experimentally determine these elasticities. These experiments have been small and of short duration, so far, and much additional data is needed.

As Wenders [13] points out, the starting point for any time-of-day pricing experiment is a workable model of demand, which shows what data needs to be measured. Such a model must be designed so that its parameters can be estimated by econometric techniques and so that it provides the information required by a peak-load pricing model like the one presented in the previous section. However, an additional degree of freedom is available here which is not usually present in econometric studies - the opportunity to design the experiment and to determine what data is to be collected and how it is to be measured. Thus, techniques from the realm of statistical design of experiments could also be brought to bear.

In addition to having a statistically tractable form, the demand model must be usable within an equilibrium model for calculating the peak-load prices. The model (12.3) of the previous section uses the consumers' surplus function as a measure of welfare. The consumers' surplus function has a number of difficulties. First, some rather stringent theoretical assumptions must be made about demand in order for this function to exist. These assumptions, called "integrability conditions" are the subject of some long-standing controversy in the economic literature. Second, econometricians are much more familiar with estimating demand functions than surplus functions, and therefore, it will be necessary to formulate estimable models for the consumers' surplus. One problem here is that since the surplus function is essentially the integral of the demand function, estimating the surplus function may require a good <u>global</u> estimate of the demand function, much more difficult to get than a good local estimate. Finally, the model (12.3) does not include prices explicitly. Formulating a model which is based on prices as decision variables may eliminate the need to work with the consumers' surplus and allow the demand functions to be used directly.

As a final point, it should be noted that while a great deal has been said here about uncertainty on the supply side, nothing has been said about uncertainty on the demand side. Yet, uncertainty about demand, both short-term and long-term, is as much a factor in planning and pricing decisions as is uncertainty about supply. Not only is the size of the peak load uncertain, but also is its hour of occurrence. Furthermore, the future growth of electricity demand is also highly uncertain. It would therefore be very useful to formulate a demand model which gives a probability distribution for demand rather than a point estimate. It seems that the analysis given in previous chapters could be carried through with uncertain demand as well, providing better estimates of expected cost and reliability. It would be necessary to extend the concept of a load duration curve to probabilistic demand. It would also be necessary to formulate a demand model including uncertainty and to determine what data would be needed in order to estimate its parameters.

Footnotes

Chapter 2 Notes

1. See Anderson [1].
2. See Massé and Gibrat [24].
3. See Fernandez and Manne [12].
4. See Gately [15, 16].
5. See Noonan and Giglio [26, 27].
6. See Anderson, op. cit, pp. 286-287.
7. A similar argument is given by Turvey in [38].
8. See Lasdon [23].

Chapter 3 Notes

1. See Geoffrion [17].
2. See Phillips *et al*. [28].
3. See Bessière [5].
4. See Beglari and Laughton [4].

Chapter 4 Notes

1. Despite its name, probabilistic simulation is not a simulation in the more usual, Monte Carlo sense, but an analytic calculation. The name is derived from the traditional use, by power engineers, of the term "simulation" to indicate the process of determining the operating costs of a generating system, by any method. This is the sense in which Anderson [1] uses the term simulation.
2. See, for example, Billinton _et al_. [6] for a brief history.
3. See Baleriaux _et al_. [3] and Booth [9]. See also Joy and Jenkins [22], Finger [13], and Vardi _et al_. [40].
4. See Schweppe _et al_. [31].
5. Private communication to the author.
6. See Beglari and Laughton [4].
7. See, for example, Billinton _et al_. [6].
8. See Telson [36].
9. _Ibid_., p. 133.
10. See, for example, Crew and Kleindorfer [10] and Turvey and Anderson [39].

Chapter 5 Notes

1. See Schweppe *et al*. [31].
2. See Finger [13, 14].

Chapter 6 Notes

1. See Jacoby [20].
2. See Joy and Jenkins [22].
3. See Finger [13].
4. See Noonan and Giglio [26, 27].
5. See Vardi, *et al*. [40].
6. See Crew and Kleindorfer [10].

Introduction to Part Three Notes

1. See Joskow [21].
2. This argument is also made by Turvey in [38].
3. See Vardi, *et al.*, [41].

Chapter 10 Notes

1. See Vardi, *et al.*, [41].
2. See Scherer, [30].
3. See Telson, [36].
4. See Vardi, *et al.*, *op. cit.*

Chapter 11 Notes

1. See Geoffrion [17, 18].
2. See Hogan [19].
3. See Shapiro [33].
4. See Geoffrion [18].

Chapter 12 Notes

1. See Crew and Kleindorfer [10].
2. See the survey by Joskow [21].
3. See, for example, Shapiro [32].
4. See, for example, Boiteaux [8], Steiner [34], and Bailey and White [2].
5. See, for example, Wenders [42], Turvey [37, 38], and Turvey and Anderson [39].
6. See Crew and Kleindorfer, op. cit.
7. Pressman [29] discusses the derivation and use of consumers' surplus functions in peak-load pricing models.
8. Shapiro [32] suggests similar approaches to general equilibrium problems of this type.
9. Shapiro, op. cit.
10. See Electric Utility Rate Design Study [11].
11. Ibid, p. 1.
12. See Taylor [35].
13. See Wenders and Taylor [43].

Bibliography

1. Anderson, D., "Models for Determining Least-Cost Investments in Electricity Supply", Bell Journal of Economics, Spring, 1972, pp. 267-299.
2. Bailey, E. E., and L. J. White, "Reversals in Peak and Off Peak Prices", Bell Journal of Economics and Management Science, Spring, 1974, pp. 75-92.
3. Baleriaux, H., E. Jamoulle, and F. Linard de Guertechin, "Simulation de l'exploitation d'un parc de machines thermiques de production d'electricité couplè à des stations de pompage", Revue E, vol. V, no. 7, 1967. pp. 225-245.
4. Beglari, F. and M. A. Laughton, "The Combined Costs Method for Optimal Economic Planning of an Electrical Power System", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-94, no. 6. November/December 1975, pp. 1935-1942.
5. Bessière, F., "The Investment 85' Model of Electricité de France", Management Science, December, 1970.
6. Billinton, R., R. J. Ringlee, and A. J. Wood, Power-System Reliability Calculations, MIT Press, 1973.
7. Billinton, R., "Bibliography on the Application of Probability Methods in Power System Reliability Evaluations", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-91, March/April, 1972, pp. 649-660.

8. Boiteaux, M., "Peak Load Pricing", The Journal of Business, vol. 33, no. 2, April, 1960, pp. 157-179.

9. Booth, R. R., "Power System Simulation Model Based on Probability Analysis", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-91, January/February 1972, pp. 62-69.

10. Crew, M. A. and P. R. Kleindorfer, "Peak Load Pricing with a Diverse Technology", The Bell Journal of Economics, Vol. 7, no. 1, Spring 1976, pp. 207-231.

11. Electric Utility Rate Design Study, Rate Design and Load Control, Electric Power Research Institute, November, 1977.

12. Fernández, G. and A. S. Manne, "A Model for Planning Investments in Generating Facilities", Internal Working Paper, Stanford University, Feburary, 1969.

13. Finger, S., "Modeling Conventional and Pumped Hydro-Electric Energy Using Booth-Baleriaux Probabilistic Simulation", M.I.T. Energy Laboratory Working Paper MIT-EL 75-009WP, August, 1975.

14. Finger, S., "Electric Utility System Generation (SYSGEN) User's Manual", Stone and Webster Engineering Corp., May, 1977.

15. Gately, D., "Investment Planning for the Electric Power Industry: An Integer Programming Approach", Research Report 7035, Department of Economics, University of Western Ontario, November 1970.

16. Gately, D., "A Mixed Integer Programming Approach to Investment Planning for the Electric Power Industry", University of Western Ontario, 1971.

17. Geoffrion, A. M., "Generalized Benders' Decomposition", Journal of Optimization Theory and Applications, vol. 10, no. 4, 1972, pp. 237-260.

18. Geoffrion, A. M., "Vector Maximal Decomposition Programming", Working Ppaer. No. 164, Western Management Science Institute, UCLA, September 1970.

19. Hogan, W., "Directional Derivatives for Extremal-Value Functions with Applications to the Completely Convex Case", Operations Research, vol. 21, no. 1, Jan.-Feb., 1973, pp. 188-209.

20. Jacoby, H. D., "Analysis of Investments in Electric Power", Harvard University, 1967.

21. Joskow, P. L. "Contributions to the Theory of Marginal Cost Pricing", The Bell Journal of Economics, Spring, 1976. pp. 197-206.

22. Joy, P. S. and R. T. Jenkins, "A Probabilistic Method for Estimating the Operating Cost of an Electric Power Generating System", Oak Ridge National Laboratory, ORN-TM-3549, October 1971.

23. Lasdon, L. S., Optimization Theory for Large Systems, MacMillan, 1970.

24. Massé, P. and R. Gibrat, "Applications of Linear Programming to Investments in the Electric Power Industry", Management Science, vol. 3, no. 2, January, 1957.

25. Moriarty, E. J., A Structural Re-Development of an Economic Environmental Generation Expansion Model, unpublished thesis, Mechanical Engineering, M.I.T., May, 1976.

26. Noonan, F. and R. J. Giglio, "A Mathematical Programming Model for Long Range Planning of Electric Power Generation", presented at the ORSA/TIMS Puerto Rico Meeting, Fall, 1974.

27. Noonan, F. and R. J. Giglio, "Planning Electric Power Generation: A Nonlinear Mixed Integer Model Employing Benders' Decomposition", Management Science, May, 1977, pp. 946-956.

28. Phillips, D., F. P. Jenkin, J. A. T. Pritchard, and K. Rybicki, "A Mathematical Model for Determining Generating Plant Mix", Third PSCC, Rome, 1969.

29. Pressman, I., "A Mathematical Formulation of the Peak-Load Pricing Problem", Bell Journal of Economics and Management Science, Autumn, 1970, pp. 304-326.

30. Scherer, C. R., "Estimating Peak and Off-Peak Marginal Cost for an Electric Power System: An Ex Ante Approach", Bell Journal of Economics, Autumn, 1976, pp. 575-600.

31. Schweppe, F. C., D. H. Marks, D. L. Farrar, J. Gruhl, M. F. Ruane, P. F. Shiers, and F. Woodruff, Jr., "Economic-Environmental System Planning", IEEE Power Engineering Society Summer Meeting, July, 1974.

32. Shapiro, J. F., "Decomposition Methods for Mathematical Programming/Economic Equilibrium Energy Planning Models", M.I.T. Operations Research Center Working Paper OR 063-77, March, 1977.

33. Shapiro, J. F., Fundamental Structures of Mathematical Programming, in preparation.

34. Steiner, P. O., "Peak Loads and Efficient Pricing", Quarterly Journal of Economics, vol. 71, no. 4, November, 1957, pp. 585-610.

35. Taylor, L. D., "The Demand for Electricity: A Survey", Bell Journal of Economics, Spring, 1975, pp. 74-110.

36. Telson, M. L., "The Economics of Reliability for Electric Generation Systems", M.I.T. Energy Laboratory Report MIT-EL-73-016, May, 1973.

37. Turvey, R., Optimal Pricing and Investment in Electricity Supply, Allen and Unwin, 1968.

38. Turvey, R., Economic Analysis and Public Enterprises, Allen and Unwin, 1971.

39. Turvey, R. and D. Anderson, Electricity Economics: Essays and Case Studies, Johns Hopkins University Press, 1977.

40. Vardi, J., J. Zahavi and B. Avi-Itztak, "The Combined Load Duration Curve and Its Derivation", <u>IEEE Transactions on Power Apparatus and Systems</u>, Vol. PAS-96, no. 3, May/June 1977, pp. 978-983.

41. Vardi, J., J. Zahavi, and B. Avi-Itzhak, "Variable Load Pricing in the Face of Loss of Load Probability", <u>Bell Journal of Economics</u>, Spring, 1977, pp. 270-288.

42. Wenders, J. T., "Peak Load Pricing in the Electric Utility Industry", <u>Bell Journal of Economics</u>, Spring, 1976, pp. 232-241.

43. Wenders, J. T., and L. D. Taylor, "Experiments in Seasonal-Time-of-Day Pricing of Electricity to Residential Users", <u>Bell Journal of Economics</u>, Autumn, 1976, pp. 531-552.

Index of Notation

Index of Notation

Symbol	Definition
$\underline{x}, \underline{x}'$	A vector $\underline{x}$ and its transpose.
$C_{jv}, \underline{C}$	Capital cost per unit of capacity of plant (j,v) and vector of C_{jv}.
$CS(\underline{Q}, \varepsilon)$	Consumers' surplus function.
$\underline{e}$	A vector of ones.
$EF_t(\underline{Y}_t)$	Expected system operating cost function in period t.
$EF^*(\underline{X}, \underline{Q}, \varepsilon)$	Minimum system operating cost.
$EG_t(\underline{Y}_t)$	Expected unserved energy function in period t.
F_{jvt}, F^{it}	Operating cost per unit of energy produced by plant (j,v) in period t (merit order index i).
$F_t(\underline{U}_t)$	Operating cost function in period t.
$f(\underline{x}, \underline{y})$	General objective function with two vector arguments.
$g(\underline{x}, \underline{y})$	General constraint function with two vector arguments.
$G(Q)$	System load duration curve.
$G(Q, Q^*)$	Step-function load duration curve.
$G_i(Q)$	Equivalent load duration curve faced by plant i.
$g_i(Q), g_i(Q; \underline{Y})$	Loss-of-Load Probability faced by plant i when load is Q.
$h_i(s)$	Loss-of-Load Probability faced by plant i in hour s.
$H_{ij}(U)$	Derivative of expected energy produced by plant i with respect to utilization level of plant j, $j \leq i$.

$H_{hvt}(I)$	Hydro-energy available from plant (h,v) in interval I.
I_t	Number of plants in the merit order of period t.
I_σ, I_σ'	Set of operative plants in outage state σ and its complement.
i	Index of plants in merit order.
j	Index for plant type.
k	Merit order index for plants in a given outage state.
k, ℓ	Index for Benders' cuts generated for the master problem.
$L(\underline{x}, \underline{\lambda})$	Lagrangian Function.
L_s	Auxiliary variable for load in hour s, Q_s.
M_t	Constraint Matrix for defining plant loading points $\underline{U}_t$.
$MC(Q_{st})$	Marginal cost associated with demand component Q_{st}.
$MC(\varepsilon_t)$	Marginal cost associated with reliability ε_t.
$m, \bar{m}$	Random variable for length of operation before a plant fails and its mean.
N_t	Constraint vector for defining peak load constraint.
n	Index of marginal plant.
p_i	Availability of plant i.
q_i	$= 1-p_i$ Forced outage rate of plant i.
Q	System load.
Q_t^*	Peak load in period t.

$\underline{Q}_t$	Vector representing load duration curve in period t.
$r, \bar{r}$	Random variable for repair time after a plant fails and its mean.
s	Index for subintervals of time and for hourly demand components.
$TC(\underline{Q}, \varepsilon)$	Total cost of building and operating a generating system; optimal value of capacity planning model.
t	Index for time period.
$U^i, \underline{U}_t$	Cumulative capacity of all plants through i in merit order (load point of plant i+1), and vector of these for period t.
U	Stands for any U^i.
$\tilde{U}^i$	Random variable for cumulative available capacity through plant i.
V_σ^i	Cumulative available capacity through plant i in outage state σ.
v	index for plant vintage (negative value implies an existing plant).
$v(\underline{y})$	Optimal value of parametric mathematical program.
$\dot{v}(\underline{y}\|\Delta\underline{y})$	Directional derivative of v at $\underline{y}$ in direction $\Delta\underline{y}$.
$W(Q)$	Integrated load duration function.
$X_{jv}, X^i, \underline{X}$	Plant (power) capacity.
$Y_{jvt}, Y^i, \underline{Y}_t$	Plant utilization level.
β_{ik}	Probability factor $q_i q_{i-1} \cdots q_{k+1}$.
$\underline{\gamma}$	Subgradient of $v(\underline{y})$.
$\hat{\gamma}_s$	Marginal operating cost associated with load in hour s, Q_s.

Γ_k	Index set of periods t in which the k^{th} trial solution $\underline{X}^k$ violates the unserved energy constraint.
$\delta^{it}_{jv}, \delta_t$	Indicator constant which converts plant indices (j,v) into merit order indices in period t and matrix which sorts plants into merit order in period t.
Δ_j	Expected difference in operating cost between plant j and next available plant in merit order.
$\Delta E_i(Y^i)$	Expected energy produced by i^{th} plant operating at level Y^i.
ε	Reliability level, loss-of-load probability or expected unserved energy.
$\underline{\eta} = [\underline{\eta}_1, \underline{\eta}_2]$	Subgradient of the constraint function $g(\underline{x},\underline{y})$.
$\theta_{i\sigma}$	Indicates whether or not plant i is available in outage state σ.
θ_s	The length of time interval s.
θ^k	Shadow price on k^{th} cost constraint in master problem.
$\lambda^i, \underline{\lambda}_t$	Dual multiplier on capacity constraint and vector of these for period t.
Λ_{ts}	Dual feasible region for LP subproblem.
$\mu^i, \underline{\mu}_t$	Dual multiplier associated with an infeasible subproblem and vector of these for period t.
ν^k	Shadow price on k^{th} feasibility constraint in master problem.
$\underline{\xi} = [\underline{\xi}_1, \underline{\xi}_2]$	Subgradient of the objective function $f(\underline{x},\underline{y})$.
ξ_σ	End point for interval of constant value for g_i.
π	Dual multiplier on demand, peak-load, or reliability constraint.

σ	Index for outage states, or intervals of constant value for g_i.
T_{ts}	Primal feasible region for LP subproblem.
τ	Time (as a continuous parameter).
ϕ_σ	Probability of outage state σ.
ψ_σ	Constant value of function g_i.
Ω	Feasible region for capacity variables $\underline{X}$.
$\|\|$	End of proof.

Biographical Note

Jeremy Alan Bloom, the author of this thesis, was born in Cleveland, Ohio, in 1951. He grew up in Levittown, Pennsylvania, and graduated from Neshaminy High School in 1969. He studied Electrical Engineering at Carnegie-Mellon University and received his Bachelor's degree in 1973.

While at M.I.T., Mr. Bloom has worked primarily on the applications of Operations Research to economic problems, particularly those concerned with energy. He worked on New England Energy Management Information Systems (NEEMIS) project and wrote his Master's thesis on "A Mathematical Model of Fuel Distribution in New England," for this project. He received his Master's degree in Operations Research in 1976. He has also worked on problems in the design and operation of computer communication networks and on a project to evaluate the impact of improvements in nuclear reactor technology. After receiving his doctorate, Mr. Bloom plans to teach and to continue his research on electric utility planning.

DATE
FILMED
-7