AD-A050 798 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 15/5
MODERN TRENDS IN LOGISTICS RESEARCH. PROCEEDINGS OF A CONFERENC--ETC(U)
1976 W H MARLOW N00014-75-C-0729
UNCLASSIFIED NL



AD A050798

DDC FILE COPY

2

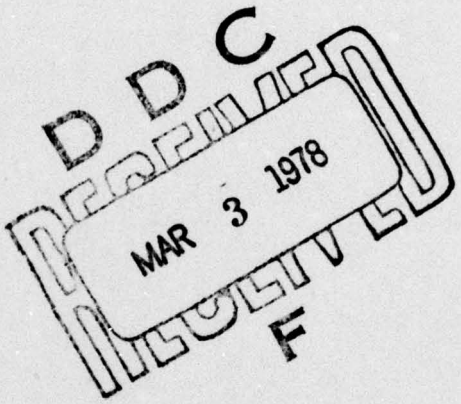
MODERN TRENDS IN LOGISTICS RESEARCH

Proceeding of a
Conference Sponsored
by the Office of
Naval Research

This volume is an outgrowth of the May 1974 Logistics Research Conference, the most extensive one on this subject ever organized, sponsored by the Office of Naval Research and The George Washington University, in cooperation with the Air Force Office of Scientific Research and the Army Research Office. The major objectives of the conference were to survey major research developments and applications since World War II, and to assess outstanding current problems and promising new research techniques. The Organizing Committee for the Logistics Research Conference consisted of Marvin Denicoff of the Office of Naval Research, and Anthony V. Fiacco, W. H. Marlow, and Henry Solomon of The George Washington University. Professor Marlow served as chairman of this committee.

The papers presented on major logistics issues by the senior logistics personnel in the Department of Defense are contained in the first part of the book. The remaining four parts contain all the invited survey papers presented for the conference. These parts and the contributors to them are as follows:

[illegible]



Approved for public release;
distribution unlimited.

MODERN TRENDS IN LOGISTICS RESEARCH

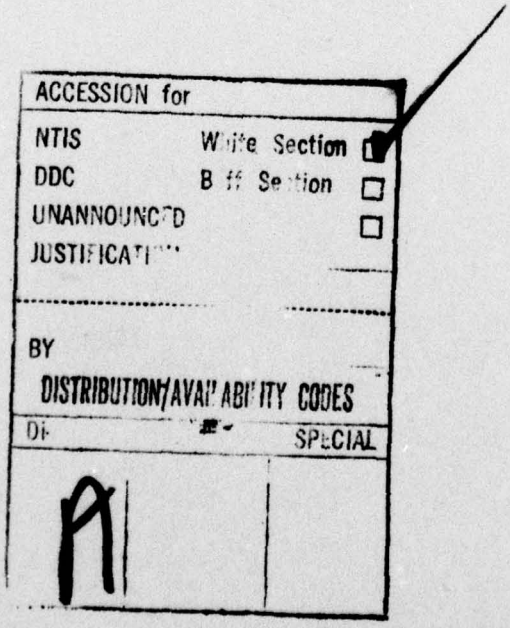
AD A050798

AD No. ____

DDC FILE COPY

PUBLISHER'S NOTE

This format is intended to reduce the cost of publishing certain works in book form and to shorten the gap between editorial preparation and final publication. The time and expense of detailed editing and composition in print have been avoided by photographing the text of this book directly from the author's typescript.



AVAILABLE COPY

(6) MODERN TRENDS IN LOGISTICS RESEARCH.
Proceedings of a Conference Held at The George Washington University.

(2)

(10) W. H. Marlow

(15) N0001475C0729

(11) 1976 (12) 478 P.

Edward S. Bres III
A. Charnes
W. W. Cooper
S. E. Elmaghraby
A. N. Elshafei
Richard L. Francis
Walter D. Gaddis
Donald P. Gaver
Murray A. Geisler
Donald Gross
Arnoldo C. Hax
Fred Kornet, Jr.
Arthur I. Mendolia
Frank Proschan
David A. Schrady
Jeremy F. Shapiro
Ronald W. Shephard
William W. Snavely
Gerald L. Thompson
S. Zacks

Edited by

W. H. Marlow

DDC
MAR 3 1978
F

The MIT Press
Cambridge, Massachusetts, and London, England

220 000

This work relates to Department of the Navy Contract N00014-75-C-0729 issued by the Office of Naval Research under Contract Authority NR 347-020. However, the content does not necessarily reflect the position or the policy of the Department of the Navy or the Government, and no official endorsement should be inferred.

The United States Government has a royalty-free, nonexclusive and irrevocable license throughout the world for Government purposes to publish, translate, reproduce, deliver, perform, dispose of, and to authorize others so to do, all or any portion of this work.

Copyright © 1976 by
The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. However, reproduction in whole or part is permitted for any purpose by the United States Government.

Printed in the United States of America

Logistics Research Conference, George Washington University, 1974.
Modern trends in logistics research.

"Sponsored by the Office of Naval Research and the George Washington University, in cooperation with the Air Force Office of Scientific Research and the Army Research Office."
1. Logistics--Congresses. I. Bres, Edward Sedley, 1947- II. Marlow, W.H., 1924- III. United States. Office of Naval Research. IV. George Washington University, Washington, D.C. V. Title.
U168.L58 1974 355.4'11 75-43617
ISBN 0-262-13122-6

CONTRIBUTORS

Edward S. Bres III is a Research Associate, Center for Cybernetic Studies, University of Texas at Austin.

A. Charnes is the University of Texas System Professor, Jesse H. Jones Professor of Biomathematics and Management Science, and Director, Center for Cybernetic Studies, University of Texas at Austin.

W. W. Cooper is University Professor of Public Policy and Management Science, School of Urban and Public Affairs, Carnegie-Mellon University.

S. E. Elmaghraby is University Professor of Operations Research and Industrial Engineering, and Director, Operations Research Program, North Carolina State University.

A. N. Elshafei is Expert, The Institute of National Planning, Nasr City, Cairo, Egypt.

Richard L. Francis is Professor, Department of Industrial and Systems Engineering, University of Florida.

Walter D. Gaddis served as Deputy Chief of Naval Operations (Logistics) from April 11, 1973 to August 1, 1975.

Donald P. Gaver is Professor and Associate Chairman for Research, Department of Operations Research and Administrative Sciences, United States Naval Postgraduate School.

Murray A. Geisler is former Director of Logistics Studies, Rand Corporation, Santa Monica, California.

Donald Gross is Professor, Department of Operations Research, George Washington University.

Arnoldo C. Hax is Associate Professor of Management Science, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

Fred Kornet, Jr. served as Deputy Chief of Staff for Logistics, Department of the Army, from January 2, 1973 to August 31, 1975.

W. H. Marlow is Professor and Chairman, Department of Operations Research, Principal Investigator of Program in Logistics, and Director, Institute for Management Science and Engineering, George Washington University.

Arthur I. Mendolia served as Assistant Secretary of Defense (Installations and Logistics) from June 21, 1973 to March 31, 1975.

Frank Proschan is Professor, Department of Statistics, Florida State University.

David A. Schrady is Chairman, Department of Operations Research and Administrative Sciences, United States Naval Postgraduate School.

Jeremy F. Shapiro is Professor of Operations Research and Management, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

Ronald W. Shephard is Professor of Engineering Science and Chairman, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

William W. Snavely served as Deputy Chief of Staff for Systems and Logistics, Department of the Air Force, from February 1, 1973 through August 31, 1975.

Gerald L. Thompson is Professor of Applied Mathematics and Industrial Administration, Carnegie-Mellon University.

S. Zacks is Professor and Chairman, Department of Mathematics and Statistics, Case Western Reserve University, and Consultant, Program in Logistics, George Washington University.

CONTENTS

PREFACE xiii

Part I: ISSUES AND PROBLEMS IN LOGISTICS 1

CHAPTER 1: MAJOR ISSUES IN LOGISTICS 3
Honorable Arthur I. Mendolia

1.1 Introduction 3
1.2 Specific Issues and Problems 4
1.3 Conclusion 9

CHAPTER 2: MAJOR ISSUES IN ARMY LOGISTICS 11
Lieutenant General Fred Kornet, Jr., USA

2.1 Introduction 11
2.2 Specific Problem Areas 11
2.3 Conclusion 20

CHAPTER 3: LOGISTICS ASPECTS OF WEAPONS RESEARCH 21
Vice Admiral Walter D. Gaddis, USN

3.1 Introduction 21
3.2 A Specific Issue 21
3.3 Conclusion 24

CHAPTER 4: MAJOR LOGISTICS PROBLEMS 25
Lieutenant General William W. Snavely, USAF

4.1 Introduction 25
4.2 Dynamic Interactive Relationships 25
4.3 Front-end Logistics 26
4.4 The Logistics Support System 29
4.5 Conclusion 32

Part II: INFORMATION PROCESSES AND SYSTEMS DESIGN 33

CHAPTER 5: THE ORGANIZATION OF INFORMATION FOR LOGISTICS DECISION-MAKING 35
Murray A. Geisler

5.1 Introduction 35
5.2 Supply Illustration 38
5.3 Maintenance Example 40
5.4 Scheduling Example 43

5.5 Data Acquisition 46
5.6 Exploitation of Data Variability 48
5.7 A Perspective 52
Selected Rand Logistics Bibliography 55

CHAPTER 6: THE DESIGN OF LARGE SCALE LOGISTICS SYSTEMS: A SURVEY AND AN APPROACH 59
Arnoldo C. Hax

6.1 Introduction 59
6.2 The Logistics Decision Process 61
6.3 The Evolution of Computer Based Logistics Support Systems 77
6.4 A Proposed Approach for Logistics System Design 83
References 89

Part III: PRODUCTION, SCHEDULING, AND FACILITY LAYOUT 97

CHAPTER 7: COST AND PRODUCTION FUNCTIONS: A SURVEY 99
Ronald W. Shephard

7.1 The Neoclassical Production Function 99
7.2 Two Stems of Development 102
7.3 The Econometric Production Function 103
7.4 Generalized Neoclassical Production Functions 106
7.5 Cost and Revenue Functions 122
7.6 Duality Between Cost (Revenue) Function and Distance Function 124
7.7 Indirect Production Functions 126
7.8 Postscript for Dynamic Models 129
References 131

CHAPTER 8: BRANCH-AND-BOUND REVISITED: A SURVEY OF BASIC CONCEPTS AND THEIR APPLICATIONS IN SCHEDULING 133
S. E. Elmaghraby and A. N. Elshafei

8.1 Preliminaries 133
8.2 Fundamentals 135
8.3 Branching 142

8.4 Bounding 160
8.5 Dominance and Feasibility 180
8.6 Miscellaneous Dicta 197
References 201

CHAPTER 9: RECENT ANALYTICAL ADVANCES IN FACILITY LAYOUT AND LOCATION: A SURVEY 206
Richard L. Francis

9.1 Introduction 206
9.2 Problem Classifications, Statements, and Results 207
References 218

Part IV: PROBABILISTIC AND STATISTICAL MODELS 221

CHAPTER 10: REVIEW OF STATISTICAL PROBLEMS AND METHODS IN LOGISTICS RESEARCH
S. Zacks 223

10.1 Introduction 223
10.2 Demand Prediction 223
10.3 Adaptive Inventory Control 229
10.4 Operational Readiness 235
10.5 Replacement When a Constant Failure Rate Precedes Wearout 237
10.6 Surveillance 241
References 243

CHAPTER 11: A SURVEY OF INVENTORY THEORY AND PRACTICE 248
Donald Gross and David A. Schrady

11.1 Introduction 248
11.2 Theoretical Survey 249
11.3 Survey of Practice 265
11.4 Conclusions 272
Bibliography 278

CHAPTER 12: PROBABILITY MODELS IN LOGISTICS 296
Donald P. Gaver

12.1 Introduction 296
12.2 A Logistics Interpretation of the

Birthday Problem 296
12.3 An Adaptive Decision Problem Involving a Markov Chain 303
12.4 Manpower and Personnel-Related Problems: Two Simple Examples 306
12.5 Repair Models 322
12.6 Conclusions 331
References 331

CHAPTER 13: RECENT RESEARCH ON CLASSES OF LIFE DISTRIBUTIONS USEFUL IN MAINTENANCE MODELING 334
Frank Proschan

13.1 Introduction 334
13.2 Replacement Policy Comparisons 337
13.3 Models for the NBU and NBUE Classes 340
13.4 Bounds for the NBU and NBUE Classes 342
13.5 Statistical Inference 344
13.6 Related Classes of Life Distributions 345
References 346

Part V: MATHEMATICAL PROGRAMMING 349

CHAPTER 14: EXTREMAL METHODS IN LOGISTICS RESEARCH: A DEVELOPMENTAL SURVEY 351
A. Charnes, W. W. Cooper, and E. Bres III

14.1 Introduction 351
14.2 Strategies for Applications 352
14.3 The Survey 357

CHAPTER 15: RECENT THEORETICAL AND COMPUTATIONAL RESULTS FOR TRANSPORTATION AND RELATED PROBLEMS 388
Gerald L. Thompson

15.1 Introduction 388
15.2 Operator Theory of Parametric Programming for Transportation Problems 389
15.3 Operator Theory of Parametric Programming for the Generalized Transportation Problem 399
15.4 The Time (Bottleneck) Transportation Problem 403
15.5 Computational Comparisons of Algorithmic Efficiency 409

15.6 Conclusions 411
References 412

CHAPTER 16: A SURVEY OF APPLICATIONS OF INTEGER AND COMBINATORIAL PROGRAMMING IN LOGISTICS 416
Jeremy F. Shapiro

16.1 Introduction 416
16.2 Discrete Programming 417
16.3 Conclusions 433
References 433

NAME INDEX 437

SUBJECT INDEX 449

PREFACE

This book is a result of the May 1974 Logistics Research Conference sponsored by the Office of Naval Research and The George Washington University, in cooperation with the Air Force Office of Scientific Research and the Army Research Office. There were six earlier logistics conferences at George Washington, most of which took place during 1949-1954 under the innovative leadership of C. B. Tompkins who was the principal founder of logistics research.

Since the early 1950s there have been several professional societies--notably, the Operations Research Society of America, The Institute for Management Sciences, the Military Operations Research Society, and the Society of Logistics Engineers--holding meetings devoted to logistics. The present conference had by far the greatest scope, as befitting the fact that it was held after more than 25 years of logistics research sponsored by many agencies at many locations. In the announcement of the conference, Marvin Denicoff of the Office of Naval Research described this era as follows.

> The near three-decade span since the close of World War II has encompassed two difficult and geographically remote wars, in Korea and Vietnam; it has been a time of tremendous growth in the number and the complexity of weapons systems; it has introduced the missile and space ages; and it has been an era of unprecedented expansion of industry and commerce. For the logistician to keep pace with these developments, entirely new tools--indeed, a whole new technological base--were urgently required. The job of meeting these requirements fell to the research community, and that community responded with a multiplicity of exciting and innovative ideas. Totally new disciplines were developed, and parts of existing bodies of theory and methodology were uniquely tailored for use in the logistics environment. Computers came into being during the period and they were directed toward the solution of logistics problems; where increased computing power alone was not sufficient to the task, the computers were augmented by research products of such emerging disciplines as operations research, statistical decision theory, and econometrics.

The main objectives of the conference were (1) to survey major research developments and applications since World War II, and (2) to assess outstanding current problems and promising new research techniques. The Organizing Committee for the Logistics Research Conference consisted of Marvin Denicoff of the Office of Naval Research, Anthony V. Fiacco, W. H. Marlow (Chairman), and Henry Solomon, of The George Washington University.

The papers presented on major logistics issues by the senior logistics representatives in the Department of Defense are contained in Part I. The remaining four parts contain all of the invited survey papers prepared for the conference and around which individual sessions were organized. It has not been possible to include any of the more than 40 contributed papers delivered at the conference.

Part I, "Issues and Problems in Logistics," consists of four authoritative and frank assessments of major logistics problems that the military services believe could benefit from research. Secretary Mendolia, in Chapter 1, presents a broad survey of specific issues and problems from the very top of the hierarchy. In Chapter 2, General Kornet discusses problem areas faced by the Army. Admiral Gaddis commences Chapter 3 with brief mention of a specific Navy concern, also cited by his Army and Air Force counterparts, for measuring readiness and relating it to resource inputs. But his main concern is with the need for proper attention to logistics in the entire weapons acquisition process. General Snavely deals with three areas for research in Chapter 4: fundamental relationships, the acquisition process, and operating logistics systems. Throughout Part I there are general issues, specific problems, descriptions of current practices, and experiences that will be of great interest to the research community.

Relatively broad topics are addressed in Part II, "Information Processes and Systems Design." In Chapter 5, Murray A. Geisler presents a number of illustrations of the use of information for decision-making in logistics and provides a survey of several topics in logistics data collection, representation, and analysis. In Chapter 6, Arnoldo C. Hax treats problems of design and implementation of logistics support systems. He supplies

considerable background, reviews what is available for use, and then proposes an approach to these problems.

Part III, "Production, Scheduling, and Facility Layout," contains three chapters. In Chapter 7, Ronald W. Shephard surveys recent developments in the theory of cost and production functions and he presents contrasts with the notion of a production function as used in econometric studies. Chapter 8, by S. E. Elmaghraby and A. N. Elshafei, surveys branch-and-bound concepts and their applications in scheduling, and inventories the basic concepts underlying the theory. These latter are presented as 14 "dicta" for branch-and-bound. In Chapter 9, Richard L. Francis surveys some recent analytical advances in facility layout and location; these apply to a wide variety of design problems occurring in logistics. In addition to their being self-contained presentations of recent advances, the chapters in Part III are further distinguished by their references to classical sources and to earlier surveys of the literature.

The four chapters in Part IV, "Probabilistic and Statistical Models," cover principal areas of research that have been closely associated with logistics research over the past 25 years. Their common objective is to assist decision-making by providing probabilistic models of the uncertainly known future. Statistical problems arise when the knowledge of the stochastic nature of these models is incomplete; S. Zacks reviews these in Chapter 10. The areas covered are: demand prediction, adaptive inventory control, operational readiness, replacement under constant failure rates, and surveillance problems. The specific field of inventory theory is one that has been a prominent component of logistics research since the very beginning. Donald Gross and David A. Schrady survey this field in Chapter 11. They highlight major developments in both theory and practice, they consider gaps between theory and practice, and they look to the future. In Chapter 12, Donald P. Gaver reviews selected probabilistic methods that are appropriate for constructing models for logistics systems. Among these methods are those from occupancy theory, Markov chains, reliability theory, and diffusion theory. Their applications are illustrated in several logistics contexts. Frank Proschan surveys

recent research on a number of classes of life distributions in Chapter 13. These classes are physically motivated for use in the study of maintenance policies, a problem area of general concern as witnessed by all four chapters in Part I.

The final Part V contains three chapters under the heading "Mathematical Programming." These deal with optimization techniques summarized as "extremal methods" in the survey by A. Charnes, W. W. Cooper, and E. Bres in Chapter 14. They provide commentary on strategies for applying extremal methods and they present a bibliography classified by methods and problems addressed in logistics research. Chapter 15 by Gerald L. Thompson treats another class of problems that has been closely identified with logistics research over the years, namely, problems of distributing an item from m origins to n destinations. He surveys recent theoretical and computational results for ordinary, generalized, and bottleneck transportation problems and illustrates a number of applications. In Chapter 16, Jeremy F. Shapiro surveys applications of integer and combinatorial programming in logistics. While these methods are relatively new, certainly as compared with those in linear programming, they have recently been under considerable development. Four types of applications are illustrated in Chapter 16: production and inventory scheduling, facilities location and distribution, pipeline systems, and routing problems.

As noted individually, the chapters in this book were prepared under various auspices. Editorial assistance was provided by Henrietta Jones, and by Bettie Taggart who was in charge of production of camera-ready copy. Final copy was typed by Olga Sylvia.

W. H. Marlow

Part I

ISSUES AND PROBLEMS IN LOGISTICS

Chapter 1

MAJOR ISSUES IN LOGISTICS

Honorable Arthur I. Mendolia
Assistant Secretary of Defense
(Installations & Logistics)

1.1 Introduction

In discussing our unsolved problems, I do not want to minimize our pride in the progress made over the years in improving logistic management in the Department of Defense. But my focus must be on the major logistic issues where progress is still needed and on problems whose solutions have so far eluded us. Our hope is that logistic research can help in finding the solutions we need. This help could be in the form of new concepts, techniques and knowledge. It could also be in the form of new and better ways of applying things we already know. When we come to look at specific problems it will be clear that both kinds of help are needed.

Before we get to the specifics, I would like to make just one point about the availability of resources that is of great importance to logistic management. In recent years the trend in Defense spending has been declining in terms of real program value, that is, in terms of the forces and support that could have been purchased by a dollar of constant value over these years. Increases in the Defense budget have been barely sufficient to cover pay and price increases for a declining force. There are such large requirements for force modernization and other needed investment that budgetary pressures to reduce logistic support costs are intense now and will become even more intense in the future. If we cannot reduce logistic support costs, we are faced with the unpalatable alternatives of deferring badly needed modernization of the forces or of reducing force levels. Both of these alternatives would involve serious risks in today's uncertain world.

In addition to unprecedented budgetary pressures, logistic managers are facing requirements for logistic readiness and responsiveness that are also unprecedented.

For example, one of the lessons in the crisis in the Middle East in October 1973 was that the effectiveness of our general purpose forces is more dependent than ever before on their capability to deploy rapidly in full readiness for combat. Effective logistic support is of course essential to this requirement for a high degree of military readiness and deployability.

The kind of Defense environment I am describing for logistic problems is not an unfamiliar one. It has the form of a problem to which optimization techniques have traditionally been applied: finding solutions that attain desired outputs (in this case logistic effectiveness) while minimizing inputs (in this case logistic support costs). Techniques and expertise that can find such solutions are needed today as never before.

1.2 Specific Issues and Problems

I will now highlight briefly some of our unsolved issues and problems in logistics. I have selected only a few--those that appear most significant from my current perspective. The papers from the military services below expand on some of these and add some others. I have arranged these topics by major logistic function or process.

Supply Management. We need to develop a better materiel distribution system for the Department of Defense, one that will improve supply responsiveness. Such a system should also achieve the most efficient mix of inventory investment costs, supply depot operating costs and transportation costs. This is a clear-cut example of an optimization problem and it must involve analyses of all relevant cost factors from a total Department of Defense perspective.

We need to improve the ways we measure supply effectiveness and the readiness of weapon systems. We also need to find ways to relate these measures of effectiveness to specific decision-making processes in supply and maintenance management, especially funding decisions. We currently have no way to calculate the effect of alternative supply and maintenance budget decisions on the level of logistic readiness. Improved analytical tools are badly needed in this area not only for better budget decisions but for better use of scarce dollar

rescurces.

A continuing problem in supply management is deciding which level in the supply system is most effective and economical for management of any given item of supply. We need better criteria to apply uniformly for decisions on whether items should be decentralized for management by local supply organizations or centrally managed by our major inventory control points.

Maintenance of Materiel. We need to find new and better ways to exploit and apply promising new concepts in maintenance policy--such concepts as "on-condition" and "condition monitoring" which have the potential of reducing maintenance costs without sacrificing readiness or safety. Related to this is the need for changes in maintenance management systems to emphasize more "sectional" repair and "piece-part" repair, as opposed to extensive and costly tear-down and overhaul.

We currently have no data systems that provide valid information on the cost of operation and maintenance of specific weapon systems through their operational lives. Such information would be extremely useful in making both logistic and program decisions through all phases of weapon system life-cycles. This is a particularly challenging problem since the avoidance of burdensome accounting and data reporting systems for the operating forces is also an important objective.

One valuable application of better weapon system support cost data would be finding the optimum time for transition from contractor support of new systems to organic support by the services. This could result in considerable savings.

Materiel Acquisition and Production. Many of our most conspicuous problems in the recent past have been in this area. While we have tackled the problem of improving acquisition management very aggressively in recent years, there is still much to be done.

Our highest priority in this area must continue to be the battle against the growing cost of weapon systems. Although new concepts and approaches to reduce acquisition costs will always be welcome, I believe our primary problem today is more a matter of practical applications than of new concepts. We need to find ways to

exploit fully and gain maximum benefit from such proven concepts and techniques as design-to-cost, value engineering, advanced manufacturing technology, component standardization, life-cycle costing, and others. These involve the process of designing and applying effective management systems throughout the acquisition cycle, from early concepts through production and deployment. Our acquisition systems must facilitate effective decisions at the right time to minimize our overall costs to acquire the needed military capability.

In the area of production, a perennial but increasingly important problem is industrial mobilization planning. We must find better analytical tools to identify the size, composition, and required readiness of the Defense industrial base for mobilization. New emphasis in military strategy requiring greater flexibility and more rapid responsiveness of our general purpose forces dictates a comparable flexibility and responsiveness of the production base. We must evaluate existing and required production capability to decide what new actions or programs are needed in peacetime to insure wartime readiness at lowest possible peacetime cost.

A related production problem is the need for a more effective Defense industrial priority system for use in emergency. The priority system must be workable in an era of dynamic changes in the marketplace involving raw material supplies, production lead times, new technology, and other factors.

Procurement Policy. This area has always been fertile in presenting complex, challenging problems to management. Foremost among today's problems is how to design more effective procurement policies and practices to assist in reducing the upward pressure on weapon system costs. We must find new and better ways to reduce long procurement lead-times, to identify and eliminate costly but dispensable contract requirements and to improve incentives for contractor cost reduction. An important contribution to reducing acquisition costs would be finding new and better incentives for contractor investment in production facilities and equipment.

Another procurement problem that is especially important in today's inflationary era is the situation of

the firm fixed-price contractor. We need to preserve the advantages of fixed-priced contract arrangements and at the same time find ways to prevent damaging losses to contractors faced with unusual inflationary pressures.

Transportation. Technological advance has significantly changed our way of doing business and has led to the need for reevaluating our transportation planning for peacetime and for mobilization. The nearly universal use of containerization today for all suitable cargo, both Defense and commercial, when considered together with the rapid decline in conventional break-bulk shipping resources, has created a new situation that requires fresh analysis.

We must identify required sealift resources in this era and determine the most economical means of obtaining them.

A related problem is to devise changes in our supply and distribution systems to preserve flexibility and responsiveness in the era of containerization. How to prevent or cope with container shortage situations is an important aspect of this problem.

Facilities Management. Although this is not an area where we would normally think of advanced research contributions, it is certainly an area with its share of management problems. As is well known, periodic announcements of base closures or reductions receive widespread public attention and concern. It is imperative that decisions affecting the base structure (retention, reduction, closure, or realignment) be made on the basis of the best possible information and analysis of relevant costs and other factors. We need better analytical tools to determine relative advantages of alternative basing arrangements considering comparative operating costs, new investment, economic impact and military effectiveness. Another important problem in facilities management is how to analyze and evaluate facilities life-cycle costs as an input to decisions on construction or the allocation of real property maintenance resources. Related to this is the problem of obtaining and using better measures of facilities condition and maintenance backlogs.

A new and very important set of logistic problems has arisen as a result of the recent energy "crisis" and the prospect of long-term tight supply conditions and high prices for energy resources. As a high priority matter we must find and exploit new concepts of operation and new technological innovations or applications to minimize energy consumption without degrading military readiness. Increased use of simulators for training is a good example of the kind of things we must emphasize, but this is only the beginning of the many changes that will be necessary in an era when energy is a scarce and costly resource. We must do no less than redefine our basic concept of efficiency to include minimal energy consumption.

Economic Impact of Defense Spending. We do not usually think of this as part of the problems of logistic management. But since Defense must obviously rely on the domestic economy for its needed resources, we cannot ignore the regional and local impact of changes in Defense programs such as procurement or base realignments.

One of our important problems in this area is how to obtain more accurate and useful data on these regional and local impacts. We need much better ways to find out where and how local economies are affected by changes in Defense procurement programs. We also need methods to assess the effect of defense build-ups on the local community infra-structure so that we can cooperate in actions to minimize adverse effects and ensure local economic support.

Utilization of Automatic Data Processing Capabilities. This is the first and most important one of a number of problems that run across the boundary lines of many of the traditional logistic disciplines. We need to obtain the most effective use of automatic data processing capabilities. Defense logistic managers have been pioneers over the past two decades in exploiting the capabilities of the computer. However, I believe we are still far from obtaining the maximum benefits possible from computer technology. One of the most promising ways to make the computer serve us better is to increase the compatibility of our numerous computer-

oriented logistic systems and, in the long run, to achieve greater systems standardization. Our long-range goal must be standardized Automatic Data Processing systems for each major logistic application, over the entire Department of Defense, while also preserving system flexibility for necessary special features of individual service or local circumstances. This goal is probably the largest single challenge to our analytical and managerial capabilities today.

Elimination of Duplication. This is another problem common to several logistic functions and once again the objective is to obtain better use of Department of Defense logistic resources. In spite of great progress over the years there are still areas where duplication exists in supply management (reparable items are an example), in materiel maintenance facilities, and in local area logistic support resources, all of which are susceptible to greater common use or consolidation. Our problem here is how to effect the necessary integration and realize the potential savings, while at the same time prevent any degradation of logistic responsiveness to individual service requirements.

Utilization of Logistic Manpower Resources. This last problem is not strictly a logistic problem, but its solution is vital to the improvement of logistic management. A large majority of the civilian employees of the Defense Department are engaged in the various logistic functions. A significant percentage of military personnel are also involved in logistics. It is obvious that success in logistic management depends on the proper training, the effective leadership and motivation, and the appropriate career satisfactions for this enormous work force. Examples of the kind of things we need for improved logistic manpower management are better ways to measure productivity and to find incentives for improvement, better ways to determine the optimum mix of military and civilian personnel in our logistic organization, and better techniques for career planning and employee development.

1.3 Conclusion

I hope that the above discussion adequately shows the

wide variety of logistic problems we consider important today. I also hope the brief treatment does not conceal the very real difficulty and complexity of most of these problems. Now I would like to emphasize two points that come to mind in reviewing such a list of unsolved logistic problems.

The first point is that in many cases it is not so much fresh knowledge or new concepts that are needed for the solution of logistic problems but better application of what we already know. I believe there is a large gap between what we can demonstrate on paper and what we can successfully implement so as to achieve results. Some may think that this is primarily the job of management, rather than of research, and I cannot disagree. However, the gap between knowledge and action is real and we must look to the research community as well as to the community of logistic managers for help in bridging this gap. I believe research can help in finding better ways to communicate new concepts and techniques to managers, better ways in selling innovations and gaining acceptance of new ways of doing business, and better ways of measuring results.

The second point is that if research is to help in the solution of the many problems I have mentioned, it must be oriented to the size and complexity of these problems; we need more "macro" as opposed to "micro" research efforts to come to grips with problems of such large magnitude. I do not wish to minimize the value of any past or present research projects, but I sincerely believe that the problems we are facing today demand the widest possible focus of research that is still consistent with the thoroughness and rigorousness of research techniques.

In conclusion, I must say that research has served logistics well in the past. In fact, it has been indispensable for logistic progress and I am sure it will continue to be so in the future.

Chapter 2

MAJOR ISSUES IN ARMY LOGISTICS

Lieutenant General Fred Kornet, Jr.
U. S. Army Deputy Chief of Staff for Logistics

2.1 Introduction

This paper deals with what I see as the principal issues and problem areas for the Army in logistics. As indicated by Mr. Mendolia just above, it is often difficult to draw the line between problems requiring improved management practices and problems requiring research advances. For many of these problems it is also hard to decide who has the primary responsibility for improving the situation. But collectively, we should be able to work on these problems and try to solve them, or at least try to minimize their effects.

2.2 Specific Problem Areas

When taken together with the issues and problems discussed above by Mr. Mendolia, the specific areas I have chosen cover a substantial range of our concerns in logistics. The first deals with overseas bases: how to react to changes, and how to store pre-positioned materiel. Next there are problems in measuring readiness, in avoiding acquisition cost overruns, and in several aspects of providing supply support. Finally, problems are cited in connection with container ships, and with maintenance processes in general. There is a common need in these problem areas to discover relationships that can lead to assessments of total effectiveness, and everywhere there is a need to exploit the potential of computers.

Overseas Bases. We are constantly pressured, by groups at home and abroad, to pull out of forward bases. There has been pressure to reduce our posture, to reduce the amount of real estate we occupy, and in effect, to reduce all of our activities at various bases around the world. Of course, this means that in addition to considering the grave strategical questions involved we must analyze the logistical impacts of changes in our forward base structure.

We need cost trade-offs for storing materiel on the scene at overseas bases, ready for use at deployment, versus retaining it at continental United States bases and assuring that it can be deployed from there. Overseas there are substantial personnel costs, as well as other types of costs, for maintaining stocks and these lead to a fairly large-size logistical support task. Alternatively, there are large transportation costs involved with home-base support. If significant tonnage must be on the scene by a specific time, as is a likely requirement in our contingency planning, then there are complicating factors. Specifically, whether it be for sealift or airlift, early requirements for moving materiel will be in competition with all the higher priority movements at the beginning. Initial requirements for movement of personnel and high priority items might well preclude satisfaction of those for spare parts and stockage equipment. In summary, the costs as well as the expected benefits are difficult to analyze in both cases: storing materiel overseas versus retaining it at continental United States bases.

There are special difficulties in assuring the serviceability of overseas inventories. Since 1968 we have been practicing to a degree the concept of support from continental United States for the European theater. We moved several of our division forces from Europe and went to a concept where we maintain and keep in Europe only their heavy pieces of equipment such as tanks, Army personnel carriers, and construction equipment. These are items that are not only heavy, but bulky, and difficult to transport as well. Our division forces have duplicate sets of these equipments at their home stations in continental United States where they also have their lighter, and more easily transportable, items.

The overseas inventory is called "pre-positioned materiel configured to unit sets" (POMCUS) stock. It represents about one billion dollars' worth of equipment. In order to protect this investment, we have been struggling for the last seven years to get the necessary dollar and program approvals to build dehumidified warehouses. Now we have some 50 out of a required 82 warehouses needed to store the bulk of this materiel.

It is not only necessary to perform maintenance on

this materiel but we also must remove items and measure their effectiveness to make sure that we are not merely holding a billion dollars' worth of rapidly deteriorating hardware. We conduct exercises periodically to determine how well we are doing: the units return to Europe, pick up their equipment, and perform their assigned mission. A few years ago, before we had perfected our warehousing concept, a team from the General Accounting Office inspected the equipment and found that it was not in good shape. As a result of their report, we are now faced with a requirement to go in once a year and inspect the equipment.

In this inspection we are required to use statistical sampling procedures and this leads to the next specific problem I would like to pose. We have been using random draws to select equipments for tests and we have been performing tests to prove whether or not the equipments would have been available for combat. Implementing these selections and establishing these proofs have turned out to be difficult in practice. We feel that there must be better ways to do this testing and we look to the research community for help.

Measurement of Unit Logistical Readiness. We must measure the readiness of all units under the "total force" or "one army" concept. We cannot do the job with the 13 active Army divisions that have been laid out for us. We are dependent upon the Reserve components to provide additional division forces, and in many cases Reserves must provide logistical support for the 13 Army combat divisions. In our method of measurement, readiness depends on (1) personnel strength on board; (2) the degree of training of those personnel; and in the logistical area (3) items that are on hand as related to those authorized; and (4) the degree of readiness of the on-hand items from a serviceability standpoint. My particular area of interest covers (3) and (4).

We are constantly striving to get a simple system: one that is workable for the unit in the field and still tells us what we need to know. We have about 1,000 items that we classify as combat essential and they need to be separated out in our reports. The list of essential items obviously varies from infantry

company to artillery battery to Transportation Corps units, and so on. For example, pieces of material-handling equipment are just as essential to a Transportation Corps unit as are rifles to an infantry rifle company. We have tried to come up with a simple reporting system, but when we simplify we also must keep some degree of logic to insure workability. The current system puts the same weights on tanks as rifles and pistols because each is a reportable item. In some respects this is unfortunate because I believe that a tank company with one or more tanks missing from its authorized 17 is less combat ready than if it had one or two of its authorized pistols missing. However, implementing a weighting system would increase the complexity over our current method at the unit level. Therefore, we allow the reporting unit subjectively to raise or lower its readiness rating, with appropriate remarks, when the statistical data do not accurately portray the actual readiness posture of the units. So there has to be some degree of logic, but at the same time the system must be simple enough so that the individual unit in the field can report quickly and accurately. It is also necessary that we be able to depend on the reports. These figures are customarily placed in fairly high places and, once having been reported, are difficult to change. So if it is once reported that a unit is in a particular condition of readiness we must be able to prove to everyone that it is truly in that condition of readiness. At the division level, with only 13 combat divisions, it is just not acceptable to be uncertain about whether or not even one or two divisions are really combat ready. Such uncertainty would make a significant difference in decision-making all the way along the line, up to and including Department of State negotiations. In summary, we need simple yet effective systems for measuring and reporting logistical readiness for Army units.

Materiel Acquisition Cost Overruns. Many things other than Department of Defense weapons systems have cost overruns or schedule overruns. They are caused by a combination of factors, not the least of which is optimism at every level of estimating. Some of that optimism is not accidental. Optimism is there because at

every level in the budget process, whether it is for building the Kennedy Center, finishing the Regional Metro System in Washington, or acquiring a new Army weapons system, the particular item in question is in competition with many others. The proponents of a particular system do not lie awake at night trying to think of ways to pad the figures so that there will never be a cost overrun. Instead, they are much more likely to lose sleep trying to determine how they can change their estimate so their project will compare favorably with its many competitors.

In addition to the ever present optimism, there is a second and interrelated factor of difficulty. In spite of all our advances in the state of the art, it is still difficult to put a hard cost estimate on something that cannot be described well at the time it must be estimated. We have gone into many aspects of parametric cost estimates that start with the very fundamentals of the old approach that Mr. McNamara started with: "All airplanes cost about the same per pound, so just tell me the weight and I'll tell you what it will cost." Nowadays, a fairly sophisticated system of parametric cost estimating is used based on individual aircraft components rather than on just the simple weight of the aircraft. Undoubtedly, there need to be many advances in existing techniques to get better estimates of what weapons systems will cost in production.

These days, we are also using something we call "design-to-cost." Personally, and this may be heretical, I do not think that this concept is significantly new or different; nor do I believe that it is going to have a major effect on our total acquisition process. The essential idea is that cost should have equal importance with other program characteristics--speed, range, weight, armor thickness, and so on--in the procurement or development process. A motor car manufacturer who knows that his competition is going to produce a $3,600 automobile will design and develop one at that price and not one at twice that figure. But there is another complication in our problem; we cannot become so engrossed in the design to unit acquisition cost that we neglect the life cycle cost. We have to be certain that low acquisition costs will not cause high costs later for repair and maintenance.

Toward this goal of not only controlling acquisition costs for weapons systems but the eventual costs for maintenance and personnel support as well, we are working on methods for life cycle costing. This is a difficult task because there are few contractors willing to sign a contract that a system will cost a specified number of dollars to maintain over, say, the next ten years. In brief, it is almost an impossible negotiating position because of the many future unknowns, and yet we must devise methods for predicting and controlling costs over entire life cycles.

Cost of Asset Visibility. This is a record-keeping problem and it is the first of several supply support problems I will discuss. In the Army alone there is an inventory of some 30 billion dollars' worth of equipment. Most of that, some 25 billion dollars, is in major or principal items and there the asset visibility is quite good, down to a certain level. We know how many of the different types of tanks, trucks, missile systems, helicopters, and so on, there are in the various locations of the world by unit, by organization, by theater, and by any way we need to know. The remaining five billion dollars' worth consists of: reparable items such as engines, transmissions, gear boxes; individual repair parts; and consumables. Here, there are about 800,000 items that the Army actually uses as compared to the 6,000 or 7,000 major or principal items. And here, with the 5 billion dollars' worth of things that are not major equipments, is where we have the asset visibility problem.

As is usual in logistics, we face a cost effectiveness problem: how much visibility do we need, and how much are we willing to pay? Computers are helpful but there are capacity limits quickly reached which cannot be exceeded. At all levels, there are limits for acquisition costs and, at lower levels, there are possible effects on the mobility or combat usefulness of units. At the highest level, we have a massive program at our National Inventory Control Points, or wholesale depots, to try to get a standardized system of reporting on items. We have six of these activities and, strangely enough, we do not yet have a standard system. But we are working toward one, and the different varieties of

hardware and software contribute greatly to our problems. We also need a standardized system for our intermediate level: the installations, posts, camps, and stations scattered around the world, and those scattered around the continental United States. At this level, we also have intermediate depots in Europe and in the Pacific. Once we get a standardized system among all of these users, we are faced with having a standardized system among the Services, and from such a system there would be many benefits.

We must in every case determine what we need to report in order to do the job, how much it will cost, and then we must find cost effectiveness relationships to help us select our reporting system. These are some of the things we are working on in the area of asset visibility.

Centralized versus Decentralized Systems. This is another supply support problem and the fundamental problem is the same as in the preceding area: how much information is needed at one place? For example, under an extreme vertical supply management concept, the Aviation Systems Command in St. Louis would have to maintain records on all of its materiel, and not on just the major items, for just about everyone everywhere in the world. Theoretically this would be possible with computers but it would take a degree of sophistication we have not yet reached. It would also take a degree of standardization or, with nonstandardization, a degree of skill in communication we have not yet achieved. On the other hand, decentralized systems by their very nature may entail duplication and they pose control problems that may lead to serious reporting difficulties. All of this brings us back to where we were at the close of the preceding area: what do we need to do the job, what will it cost, and how are we to find workable relationships to help us decide on our best course of action?

Vertical versus Horizontal Stock Funding. The next question is: how do we fund for all of this? We have an ongoing and, it seems to me, continuous dialogue on the relative merits of these two schemes. In its simplest terms, vertical stock funding corresponds to

having just one checkbook to handle all accounts at the wholesale level for items to be issued for use. In horizontal stock funding, items are sold from one level to the next until they come under the retail stock fund, and then they go to the customers. This latter system works well in our command structure because it gives our individual commanders the responsibility and authority over the stocks in their own theaters or areas. The advantages of changing to a vertical system would be that there would be just one financial transaction and it may be possible to move items more freely. But in our review thus far we have not found that these advantages would outweigh the disadvantages, especially when we consider the cost of changing systems. But we continue to study the possibilities.

Our automated systems are factors for consideration here just as they are for all of the problem areas I have cited. Computer systems are surely here to stay and we are relying more and more on their use. They allow us to handle many more supply transactions and we have a much better visibility for what is on hand than ever before. We are able to monitor and manage our maintenance systems better than before. We have standard reporting systems for keeping track of what is in a port, what is leaving, what is due in, and so on.

Direct Support System. Automation has made possible, or at least certainly helped, what we call the direct support system. Since 1971, we have been working on a new concept for our support or supply procedures which historically were based on a system of echeloning supplies. Previously, large quantities would be stocked at a large depot in continental United States which would receive output from production. Overseas depots would requisition on the continental United States depot and maintain a level of 15 days-or-so stocks to support units in their theaters. Finally, the consumers carried enough of their own stocks to keep them going in a wartime situation for a few days. This is a good, reliable, safe, workable system but unfortunately it ties up a lot of dollars in inventory in the many locations.

Our new direct support system cuts out some of the middlemen. Requisitions for high priority items are

handled in much the same way as before: the closest supply support point fills the requisition. But requisitions to keep bins full go directly from the support unit to the continental United States. The order is filled at the depot, it goes to a containerization/standardization point, it is flown or shipped to the unit, and it is offloaded at that unit. Success, once again, depends on a combination of speed and accuracy. We are gradually achieving speed but accuracy is proving to be more of a problem. All the units in Europe, and most in the Pacific, are on this new system. We see this as the only way to reduce inventory so that we can live within the dollars available in the budget for logistics.

The Container Ship Revolution. Containers have completely revolutionized our way of transporting supplies. Ten years ago, 3% of our ships were container ships and now they represent about 35%. If we exclude ammunition--because we do not yet have containers for ammunition--shipments to Europe are as high as 80% containerized. We have about 50% of the ships that we had 10 years ago and these can do the same job as could twice their number 10 years ago. There are two reasons: faster sailing times and improved capabilities provided by containers.

There are disadvantages. Containers work just jimdandy in peacetime in a port with all the tremendous cranes and material-handling equipment to take them off the ship and onto tractor trailers to roll down the road. We find it less workable trying to offload from a rolling ship anchored off a beach somewhere. We have run a number of tests, in conjunction with the Navy, trying various means, including helicopters and even balloons. Our object is to come up with a solution that is completely workable in wartime. Not that everything in wartime must go over the beach, because we would still attempt and expect to use many ports that are in existence today. But we need to be ready, and to have the capability to go over the beach.

We have also learned that there are problems with containers which, after all, are nothing more than small shipping spaces. While we have used operations research and systems analysis to be sure that the ships

are not all at one place and the materiel somewhere else, we now must do the same on a much broader scale with the containers. In early 1974 we found that while we were busy shipping containers into Europe, they were not moving anywhere. And for a while we had them all in Europe and that was not helping us in various ports on the Eastern coast. There is also the problem that in certain economic situations the containers become very attractive to commercial shippers. When the dollar was devalued in both Europe and Japan, persons in those countries suddenly placed orders to buy things they had not been buying and they asked for delivery by containers. We were faced with a container shortage for the first time in years. We took a number of our own containers from storage, threw them into the breach, and got over that hurdle.

In summary, containers pose a series of problems that have to be examined. All the problems are not solved as soon as materiel is put in containers and shipped. Nevertheless, the advantages outweigh anything we have found in the way of disadvantages.

Maintenance. This is another problem area that we have in the Army and I believe it is a problem in all the Services. In brief, our preventive maintenance program has caused difficulties, some of which it was designed to prevent. The periodic tearing down of equipment to prevent maintenance may be as contradictory as it sounds. We are looking for ways to reduce preventive maintenance requirements. We are also looking for new incentives to attract the kinds of personnel we need and to provide a program that will improve the quality of maintenance.

2.3 Conclusion

I have reached the end, not because I have presented all of my problems, but because I have briefly covered the areas I selected as significant and representative. I hope that I have not conveyed the idea that life is nothing but a problem. We do lick some of them once in a while. But I did not take this opportunity to go over some of our substantial accomplishments because I think it is more important that I describe some of the things that we are working on and on which we would appreciate help from the research community.

Chapter 3

LOGISTICS ASPECTS OF WEAPONS RESEARCH

Vice Admiral Walter D. Gaddis, USN
Deputy Chief of Naval Operations (Logistics)

3.1 Introduction

The Navy has undertaken considerable logistics research in the past and there have been substantial benefits, for example, in areas of inventory policy, transportation, and modeling. We are also engaged in logistics research today, for example, in computer internetting, decision analysis, and procurement costing methods. Actually, I could provide a list of specific problems for research, some broad and some narrow, facing the Navy in the logistics field. An example is our need for a simple, usable definition of material readiness of Naval forces, a means of measuring it, and some perfectly definite input-output relationships. We need to be able to link resource inputs, and this means money, to any of the numerous potential outputs, and these mean military applications. We need to be able to predict not only how much the readiness measure will change, but also when it will change, as a result of changes in inputs. Finally, this readiness measure must be usable by horny-handed military managers.

But instead of dwelling on the past, or summarizing the present, or challenging the research community with a list, I have decided to address a specific issue, namely, the logistics aspects of weapons research. I will make clear what I mean by this term but first I would like to stress that it identifies an issue which the Services and the research community absolutely must face together to enhance and assure the strength of our defense capability.

3.2 A Specific Issue

One of my main intentions is to impart the feelings that any commander has when he is carrying out his mission ashore or afloat. Like my Service counterparts, General Kornet and General Snavely, my role as Deputy CNO for Logistics is to advise the Chief of Naval Operations on all logistics matters, and to plan and assure provision

of logistics support, for Navy units afloat and ashore. And the units of our multi-mission Navy consist of some 510 ships, 6200 aircraft, and better than 1500 shore activities. The emphasis is on the ships and the aircraft, and this emphasis is driven with a degree of urgency dictated by the fact that the deep-water navy of the Soviet fleet has, in 1974, surpassed the U. S. fleet in numbers of ships, in the low average age of these ships, and in the sophistication of some weapons systems.

I will introduce my concern about the logistics aspects of weapons research with an example. When the integrated logistics support concept was being developed, and the logisticians had envisioned that the managers would know what assets they had, how many were available, where they were located, and what systems and weapons each of the parts supported. It did not take long, however, to realize that such a data bank would be too cumbersome and too costly. But this did not stop integrated logistics support concept development, which indeed is valid and vital in today's world of reality in life cycle costing, in getting the best return from the ever shrinking defense dollar, and in competing with public and private sectors for industrial capacity, which is now approaching saturation. Instead, we find ourselves approaching this logistics data base on an incremental basis through individual weapon systems developments and by increased sophistication in computer technology and management techniques. There is, however, a distinct danger in this sophistication that we must keep keenly in the forefront: the human element in weapon and systems operation and maintenance. This is not to say that our weapons will be technologically inferior to the weapons of any potential adversary. Indeed, they must be equal to or better than those of a potential enemy. But, in my role as a logistician, and in my past duties as a commander, a common problem continually surfaces. This problem, and the issue, is reliability and maintainability versus an operational capability. And this is what I believe must be addressed through proper attention to the logistics aspects of weapons research.

In this age of the race for technical superiority, we are producing weapons systems which are truly

advanced and which give us a quantum increase in operational capability. This is due to the efforts of the research community in keeping abreast of, and applying, modern technology. However, if the weapon system is not operational every time it is needed, then all of our sophistication and technological superiority go down the drain. What I am simply saying is that the research community may offer a superb capability but if this capability cannot be maintained at an acceptable degree of reliability, we are gambling that a gun will not fire when it must fire, or that a vital system such as a radar will not function when needed.

I think a common failing that exists is that in producing capability in a piece of equipment, the fact that it must be maintained in an operational environment by the people who will be available is often overlooked. The degree of sophistication must not exceed the level of technical expertise that is reflected in our bluejackets today. The average sailor of today has a ninth grade reading and comprehension level, but with technical training he will have achieved a high school comprehension level. Although the Navy invests sometimes up to two years of technical schools in these men, their ability to diagnose and repair systems designed by PhDs, for operation by PhDs, is lacking.

This can be illustrated by another example. The SPS40 air search radar was introduced into the fleet in 1961. It could locate and track aircraft at 45,000 feet and 250 miles and we needed that type of performance badly. But it was a nightmare when it came to keeping it operating. I had command of an amphibious group with five SPS40s in ships of the force and these were our only air search radars when we were operating without a destroyer screen, which was most of the time. During a major portion of this time we had only one of these radars operating and 60 percent of the time it was the one on my flagship because I leaned on the skipper so heavily. Now 14 years later, we still have problems maintaining this radar. We have gone through three major modifications, and we have spent a considerable sum in doing it. The SPS40 was pressed into service to fill a need which still exists; under the pressure of urgency, not enough time was given to assuring its maintainability. In retrospect, an extra

year or two in its development to make sure it was maintainable would have been a wise and economical move.

The problem is that to provide a capability a system must be reliable and maintainable. For those accepted periods of downtime for preventive maintenance or system malfunction, the system must be designed to be maintained by the level of technological competence of those operating the equipment. The task ahead, then, is to design capabilities which are superior, to allow high reliability factors, and to assure that maintainability is within the technical expertise of those faced with its operation. There is an adage within the Navy which states: "Nothing is sailor-proof." This is most appropriate, since equipment which functions extremely well during evaluation under the care of highly skilled technicians does not always function quite that well under the rigors of its operational environment. Certainly then the logistics aspects of research must be concerned with the actual details of the operational environment.

3.3 Conclusion

I realize that I have dealt with a broad generality, and I do not mean to imply that all systems now in use or contemplated for the future are to be valued in low esteem. However, there are systems or components of systems which have been technically superior when introduced, but which have failed miserably because they were in a down status more than they were in an operational status. Complexity and sophistication are impressive and invoke wonder in certain circles, but when we are on the line, working equipment is a bit more reassuring. We can be sure with the shrinking defense dollar and accelerated cost of needed system capabilities we must give the taxpayer the most firepower and system performance for his tax dollar that we can. The name of the game is a ready and strong line of defense to preserve our nation's peace and to protect it in a time of crisis. The research community can contribute by developing systems and weapons that are reliable, maintainable, and supportable and these, the logistics aspects of weapons research, are what I consider the <u>real</u> problems of research and development.

Chapter 4

MAJOR LOGISTICS PROBLEMS

Lieutenant General William W. Snavely, USAF
Deputy Chief of Staff (Systems and Logistics)

4.1 Introduction

My objective is to present a few ideas on major logistics problems confronting the Air Force to stimulate thinking on the kind that can be attacked in research. I shall consider three main areas of interest. First, there are the dynamic interactive relationships in logistics systems. Second, there is the acquisition process which brings material into being and into the logistics system; I prefer to call this front-end logistics. Third, there are the processes of the logistics support system itself; these are the processes that control management and replenishment of the logistics system such as determination of requirements, distribution, maintenance, service engineering, procurement, and their associated individual management processes.

4.2 Dynamic Interactive Relationships

Logistics is a vast subject composed of the many elements of determining requirements, acquiring what is needed, and distributing or otherwise assuring that things are taken care of. Each element represents a whole family of problems for study. However, what must be understood is that the elements are highly interactive. They must work together at a variety of levels throughout the logistics spectrum. This is fairly well understood at the higher, or macro, levels and yet there is extreme difficulty in clearly establishing, or quantifying, cause and effect relationships between parts. Herein lie many of the difficulties experienced by logistics managers in the areas discussed below.

To solve this cause and effect problem, logistics managers and researchers must achieve a good understanding and a valid description of what impacts the separate logistics processes have on one another. This is essential because management effort and organization are functionally oriented and focused on optimizing performance of one process or segment of the logistics

system. Even policy development and implementation tend to be along functional lines. The result is a "disconnect" that hampers our ability to perceive cause and effect relationships and a full understanding of the factors within the individual processes that determine costs and establish effectiveness. This leads to the other areas I will discuss: areas that are high resource consumers or which determine other costs.

4.3 Front-end Logistics

As briefly suggested above, front-end logistics deals with the interaction and impact of the following on the logistics system: (1) elements of design such as reliability and maintainability; (2) concepts of operation and maintenance; and, (3) processes such as procurement. To set the stage I will comment on some of the actions taken by the Air Force to improve front-end logistics.

We have recognized for quite awhile that if life cycle costs are to be reduced, a major part of this reduction must occur at the time of system definition and development. We have established logistics organizations within the Program Management Office in effect to be responsible for logistics throughout the acquisition period. We have about 40 of these today. It is their job to assure the influence of logistics on systems design, to design the support system, and to plan the support for the system designed. In addition, the Air Logistics Command has established a Deputy Chief of Staff for Acquisition under a major general to bring together in a cohesive way all of that Command's efforts to interface logistics considerations during the acquisition process. This is a sizeable organization of some 250 people.

Progress has been made on focusing attention on logistics effects during systems design which has resulted in significant improvements in the design of our newer systems such as the F-15, A-10, B-1, and AWACS. Many of these systems have three times as many quick disconnects, four times the number of access doors, one-third fewer hydraulic filters, and one-third the number of instruments of some of our latest systems presently in the inventory. As a result these systems will require 15 percent fewer maintenance personnel and have

turn-around times almost half as long as their older counterparts.

To take advantage of knowledge gained from supporting systems we have in being, we have placed considerable emphasis on the development and use of "increased reliability of operational systems" (IROS). This is a program: (1) to collect maintenance or technical deficiencies observed in the field and depot; (2) to record systematically these data by weapon system and component; and (3) to rank the components within the system on the basis of cost to support, impact on safety, and impact on mission effectiveness. We have found that typically 1% of the items account for 30% of our costs, and that 50% of our costs are caused by 7% of the items involved. The IROS data are compiled into handbooks together with the modification history of the components and these are being made available to designers and technicians for use in the design of new systems, engineering changes, and modifications.

However, we have only begun to scratch the surface in use of IROS data. Additional analysis and study are required to exploit fully the potential represented. We anticipate payoffs such as the development of logistics forecasting techniques for use on major systems for determining the impact of changes in design, for forecasting future production and support costs, and for determining whether further improvement in reliability and maintainability is preferable to increasing the number of spares or increasing maintenance man-hours. These are only a few of the uses that can be made of IROS data. The whole area of achieving the potential promised by these data should be particularly attractive to researchers. There is a large data source and there are people knowledgeable about the data being reported.

The use of life cycle cost models to identify logistics effects is another area in which we have made progress. These models have been, and are being, used in the development of new systems and equipment. Though we have achieved some notable successes, such as the use of the "acquisition based on logistics effects" (ABLE) model in the A-10 program, much of our work has been characterized by large amounts of effort for small achievements. Our program here is not the development of new models, but ways of exploiting the models we

already have. In this area there is ample challenge for both the research and management communities to make workable the models we have now. Actually, we need to make them work with data that are also available. Much work also needs to be done in translating the logistics effects noted by our models into design-tradeoff choices, not just for the manager, but also for the design engineer to use on his drawing board. We have a long way to go in assuring that the design engineer understands the effects of design choice on total systems cost as well as on performance.

Another front-end logistics process that interacts with several functional areas is provisioning; this accounts for three to four hundred million dollars of spending per year. It interacts with such things as the requirements determination process, procurement, and pricing activities of both the contractor and the contractor production planning activities. Provisioning is dedicated to putting into the logistics system the parts and aerospace ground equipment required to support initial operational use of a new weapon system. The interface between the ordering of spares and the production-purchase planning of the prime contractor is not well understood and therefore methods to handle effectively these interfaces are not present in our procedures. Based on initial studies, it appears that a cost avoidance of two to six times the unit cost now being paid for many of our provisioning spares is achievable through accommodation to, and understanding of, these interfaces. This represents a most fruitful area for definition of cause and effect relationships of the interacting processes involved, and for further research.

There is one more area of interest related to the front-end logistics process, namely, that of "war readiness spares kits" (WRSK). These are kits provided our combat units to permit them to deploy away from their home bases. The objective is to provide a period of operational support up to 30 days, utilizing a "remove and replace" concept until resupply can be established and a maintenance capability can be developed at the operating site. Some of our WRSKs cost well over 50 million dollars each and our current inventory amounts to around 428 million dollars. Under our

current WRSK authorization and maintenance criteria we would have to spend another 327 million dollars on spares to meet our WRSK objectives by the end of fiscal year 1976. This would result in a 755 million dollar WRSK inventory--a considerable amount of resources. Therefore, we have undertaken several studies in this area. One of these is to test the effectiveness of current A-7, F-4, and RF-4 WRSKs in support of planned wartime activity (Saber Readiness Delta). Another has the objective of minimizing the cost of WRSKs needed to support planned wartime activity by determining the most effective and efficient WRSK composition. Some of the questions we are trying to find answers to are the following. (1) Could early deployment of intermediate maintenance (added repair capability) reduce WRSK size? (2) Would this early deployment of maintenance resources compound resupply problems of deployed units? (3) Would enhanced air resupply with its resultant resupply times preclude the need for deployment of large intermediate maintenance forces with their own attendant support requirements, that is, food, clothing, shelter, and other personal needs? (4) Would reduced air resupply times provide a cost effective tradeoff in lieu of deployed intermediate maintenance without loss of combat effectiveness? These questions and many other matters relating to WRSKs are important and require our attention and study. The Air Force would welcome the efforts of the research community to contribute to their resolution.

4.4 The Logistics Support System

The problems here relate to the processes that control the management and replenishment of the logistics system. The primary question involving the logistics system concerns readiness. Is the logistics system capable of supporting sustained combat operation? Readiness determinations within the Air Force today are focused on the combat unit level. They are geared to answer such questions as the following two. Is that unit equipped, manned, and trained to meet its wartime tasks assignments? Does it have on hand those supplies, munitions, and specialized items to meet its combat commitments? Underlying this assessment is the assumption that the logistics system can sustain these

operations. This brings up such questions as the following four. What is our stock availability and its status worldwide? What is our depot maintenance capability? What is the availability of bits and pieces to support accelerated maintenance lines? What is our transportation and associated capability to prepare and move materiel through the system? Today we have individual performance indicators such as "not operationally ready supply" (NORS), and back order rates, and pipeline times, to assist us in making these assessments. The problem here is in determining what these indicators really mean: if they reflect a negative trend does it mean that the logistics system is not ready to meet its commitments, or are they merely reflecting a transitory condition? More importantly, what needs to be done to alter these trends? What is the cause and effect relationship amongst the various functions they monitor? Information particularly concerning cause and effect is not available today, and it is badly needed because the answers to these questions have grave impacts on planning, programming, and budgeting decisions and on management assessment of significance and determination of corrective actions to be taken.

To answer the above questions and to find the cause and effect relationships require a great deal of study and the development at least of what I would call a "loose" or informal model. We are undertaking such a study in attempting to develop a logistics readiness model. Its objective is to bring together various information systems now available that cover each segment of the logistics system and to see if these segments can be related in a meaningful way. We hope in so doing we will have the capability to assess by model-series of weapon systems the ability of the logistics system to meet assigned war missions and to determine the impact of unprogrammed flying hour surges, budget reductions, or materiel drawdowns to support national policy. And finally, we hope to be able to evaluate alternative resource allocation strategies in terms of wartime requirements. Any assistance the research community can give in attaining our goal would be appreciated.

Another area of great interest that needs continued study is maintenance. It accounts for very large costs. Nearly half the people in logistics are involved in

maintenance and logistics accounts for about one-third of the people in the Air Force. Consequently, there is a high payoff in concentrating on examining maintenance processes. We in the Air Force have done an extensive amount of study and we have taken several actions to improve our management techniques through the use of automation and standardization programs such as "maintenance management information and control system" (MMICS) and "rivet rally." Yet much needs to be done to examine the processes themselves. We need improved understanding and definition of what should be accomplished at the various levels of maintenance, particularly at activities involving depot and intermediate base maintenance actions. Some of our recent studies indicate a redundancy of effort between these maintenance levels. The questions, of course, are the following. Should these activities be performed at both levels, and what should be the division of work between the various levels? What should the inspection interval for equipment be and what should the work content be? This is another area for fruitful research because data are available and there are people who are knowledgeable in various aspects of the maintenance system.

Another area of depot maintenance that requires study is the question of the sizing of our depot maintenance facilities. To date the push has been to consolidate depot facilities under the assumption that we can economize through consolidation. In essence we are in search of that sizing of our industrial facilities that will maximize our economies of scale while providing the support required. The major problem here is that research literature abounds with the theory of economies of scale, but little has been done in actual studies concerning real-world production facilities. And what has been done is concentrated on looking at manufacturing operations where the processes are well-defined and repetitive in nature. The maintenance activities of our depots are different in nature. They are determined by inspection requirements or by what needs to be done to repair a piece of equipment or weapon system to return it to full service. Therefore, the extent of the work involved and parts required can vary greatly from one situation to another. For example, the limited studies we have conducted to date indicate

that considerable variability exists in the amount of man-hours required to repair the same type of components going through the same repair line. In researching this area, the questions that must be answered are the following. What is the optimum size repair facility? What impact does the range of items have on the sizing of a facility? What are the factors that must be considered in sizing a depot maintenance facility? This list could be continued but the size and scope of the effort involved should be apparent by now. I hope the logistics research community will undertake this effort for the payoff promises to be large in reducing the 1.4 billion dollar yearly expenditures for our depot production facilities.

4.5 Conclusion

I realize that my discussion has been rather broad and generally encompassing and yet I have only been able to touch on some of the problems we are dealing with in Air Force Logistics. I hope I have achieved my purpose of providing a frank discussion of some of our important problems and that I will stimulate interest in accomplishing research in these areas. The research community is a fundamental ingredient in our efforts to solve these problems and reduce logistics support costs. We most certainly need help in selling and implementing innovations, in measuring the results of these efforts, and more importantly, in narrowing the gap between research findings and practical application of these findings.

PART II

INFORMATION PROCESSES AND SYSTEMS DESIGN

Chapter 5

THE ORGANIZATION OF INFORMATION FOR LOGISTICS DECISION-MAKING*

Murray A. Geisler
The Rand Corporation

5.1 Introduction

This survey was prepared for the session on logistics data collection, representation, and analysis, which covers a wide subject area. Therefore, the material that can be presented must be narrowed, or at least focused. To help in this endeavor, the general theme selected for this paper is the organization of information for logistics decision-making. Even this focus is quite broad, and the paper will be specialized further in terms of the subject matter in logistics that is stressed. The use of information in logistics decision-making is itself an evolving one, which has its own foundations in many disciplines. Some of the early work in logistics draws heavily from statistics, particularly of the descriptive kind, such as portraying frequency distributions and endeavoring to determine the probability law that seems to describe the demand for spare parts. Such descriptive work provided very useful insights into the nature of logistics problems. For one thing, it led to the general notion that demands for logistics resources are highly variable in both frequency and quantity. This early work also indicated that the costs of logistics resources vary considerably.

Other insights helpful in early logistics research came from economics, with efforts to apply the theory of the firm to logistics situations. These concepts dealt particularly with notions of marginal analysis and marginal productivity. These generalized concepts were helpful in the development of inventory theory, and in the recognition that logistics resources can be

*This chapter has been published (in slightly revised form) in M. A. Geisler (ed.)(1975). Logistics. North Holland/TIMS Studies in the Management Sciences 1 35-51.

traded off against one another in order to attain specified objectives at lower cost.

Of course, the techniques of operations research have had their major role in logistics, and these techniques make heavy demands upon data systems and the processing of such data with the assistance of computers. Much of this discussion will center around the period during which operations research helped to develop our understanding and progress in logistics.

The survey thus has the following structure. It contains a number of illustrations from logistics on how information has evolved in its role for decision-making. The illustrations cover a few of the earlier uses, with some indication of how these uses have evolved into more complex requirements. The illustrations are then followed by an overall perspective on the current state of the field, with some discussion of where the challenges now exist.

The overall survey is supposed to cover the subject of data collection. This is a very large topic, and it involves methodological and even technological issues. As will be seen, this subject will be treated in the context of the specific examples rather than trying to treat the subject in its entirety. However, a few comments might be made on the data collection area to indicate its content. Data, and hopefully good data, are the essence of logistics management. So much hinges on inventory knowledge, be it of people or things, and the transactions that occur with such logistics resources. Earliest progress on data collection both in inventory and transactions was made in supply, maybe because the counting and control over inanimate objects is a much easier operation than the counting and control of people, such as in maintenance.

In a similar manner, the first uses of computers were made in the supply area, and as a whole this area has progressed further than others, not only in the data collection process, but in the general level of logistics management. However, even in supply, there are conceptual and definitional issues that affect the data collected. Thus, there is the perennial question of the meaning of demand. Typically, the data collected identify "issues," which in concept is different from "demand." Efforts are made in some data collection

systems to obtain demand data, usually in addition to issue data, but these lead to complex coding and programming problems, with consequent costs and arbitrariness in the data obtained.

The reporting of maintenance man-hour data, which is fundamental to maintenance management, is affected by various biases. Since these data are used to evaluate and justify maintenance manpower needs, there is the persistent feeling, with some evidence, that such numbers tend to be inflated in order to report that people are kept busy. Aside from having its effect on determining realistic manpower requirements, biased data can have their unfortunate effect on other logistics policy issues, such as where maintenance functions can be economically located.

This interaction between data collection and reporting, and institutional behavior, means that logistics data cannot be treated as abstract numbers of objective information. The user of such data must be familiar with the institution and the incentives that affect those who report the data, and this knowledge must be an integral part of the interpretation that is given to results obtained from the use of such data.

Data collection also deals with such subjects as forms, consistency checks, and in certain areas with the possible use of new technology in collecting data. Some of the problems created by the availability of this new technology will be raised subsequently in this survey, but there are also certain logistics areas where it is clear that the existence of new capabilities could be most beneficial. The reference in this case is the supply distribution function in logistics. Studies of supply pipeline data indicate that the addition of faster means of transportation has not necessarily led to much reduction in pipeline times. The reasons for this result seem to lie in the delays incurred as the resupply action passes from one phase to another. Often these delays are associated with the paperwork involved and the manual actions required by the people in the process. Technology involving special encoding and recording devices has been tried to facilitate the recording of the paperwork and the movement of the cargo through the successive stages of the pipeline, with some success. It does look like a promising area, but it is

just one of many in which technology can contribute to the data acquisition area and at the same time facilitate the logistics process.

The remainder of this survey will concentrate on the other topics in the area: representation and analysis, with the focus on the theme of the organization of information for logistics decision-making. This writer's experience in logistics in the past 20 years has been almost exclusively confined to Air Force studies. Consequently, the illustrations deal with the Air Force logistics system and aircraft. However, it is believed that the nature of the problems discussed, the results developed, and the current state of the field are quite general in their relevance to the other services and to the theory of logistics, if such can be assumed to exist as an entity.

5.2 Supply Illustration

In the early days of supply and inventory analysis, a basic insight was developed with respect to spare parts used on such weapon systems as aircraft. This knowledge was gained through a special data collection system used by the Air Force in the early 1950s to help provide insights into the inventory management problems then being encountered with the B-47 aircraft. From these special data, frequency distributions of demand over substantial periods of time, such as a year, showed that relatively few items were demanded, and that of those that were demanded a very few accounted for the bulk of the demand generated. This experience varied from earlier supply behavior, in which for categories such as food, fuel, clothing, and so on, the demand or usage tended to be comparatively stable, so that the terminology "days of supply" was standard in supply manuals and statements about supply requirements. However, such terminology does not fit very well the sporadic and variable demand of spare parts. Only in recent years have some of the official publications dropped the notion of days of supply in dealing with spare parts stockage.

This aircraft demand analysis led to the use of probability models for setting supply policies. Efforts were made to fit probability models to observed demand behavior. The Poisson distribution, with its single parameter, was an early favorite, followed later

by the negative binomial, compound Poisson, and so on. These more complex demand models were intended to reflect uncertainty about the demand parameters as well as the variability in demand itself related to the stochastic nature of the process.

These probabilistic insights led to the notion that the inventory problem was one of providing a given amount of supply protection at minimum supply or system cost or maximizing the supply protection for a given supply budget. Such formulations were made in the early 1950s. The so-called flyaway kits or allowance tables were early favorite models making applications of these supply insights. One continuing difficulty with these models--and it persists today--is the so-called range versus depth problem. Since so many of the items have zero demand in any base period, what assumption should be made about the demand parameter for such items? The use of a compound Poisson or negative binomial distribution helps to provide a model to reflect this dilemma, and it seems to work relatively well. This problem also helps to highlight the system point of view that needs to be taken with supply or inventory policy.

Because of the high uncertainty attached to the demand parameters of most spare parts or line items, the inventory manager must view the supply management in a system context. He cannot guarantee inventory availability item by item under budget and space constraints, but he can only aspire to a level of supply effectiveness for the given weapon system or group of items. Thus, the expectation is that certain of the zero demand items in one year will have a demand in the next year. Since inventory may not be on hand for all zero demand items for most inventory systems, there will be stockouts, but for the system taken as a whole a given percent of demands or some other related index of supply effectiveness is expected to be satisfied.

The supply data analysis problem has also dealt with the question of an adequate performance measure for supply. Many have been developed and used: fill rate, stock rate, backorder days, and so on. In one way or another, these measures are themselves proxies for more comprehensive measures of supply performance, such as NORS (not operationally ready supply). The question of

which of these measures to use tends to be determined by validity, custom, ease of computation, and the particular mathematical formulation involved. Certain of the measures have mathematical properties that facilitate the use of optimization techniques, and others seem to provide an inventory policy that is closer to the objective function deemed most realistic for supply management.

Aside from probabilistic models to reflect demand behavior, efforts have been made to employ demand forecasting techniques, especially various kinds of smoothing formulas. The success with these techniques in forecasting demand for aircraft spare parts has been spotty, probably because of the generally low demands for such spare parts. The particular smoothing formula does not seem to make that much difference in the effectiveness of a supply policy. The recognition that the inventory process needs to be considered in a probabilistic context, and that the problem is one of resource allocation under uncertainty, seems to contribute much more technically and operationally to inventory management systems involving stochastic demands.

The design of computerized inventory systems has also reflected the knowledge that demand for many items seems to be low and erratic. This insight has led to the notion of having current centralized inventory information for control purposes, so that assets can be responsively moved around the system, from one location to another, as demand causes changes in inventory levels or as new assets enter the system. This kind of centralized management seems to be especially well suited to so-called high-cost items because of the limited quantity of such items that are procured for inventory purposes. Such control systems must deal not only with the distribution problem, but with the component repair problem as well. This situation requires control systems with multi-echelon capabilities, since current logistics structures permit both stockage and repair at several echelons.

5.3 Maintenance Example

A major insight in maintenance developed a number of years ago in the late 1950s when it was clear that the data available for aircraft maintenance management lacked some critical data elements that are necessary to

obtain better control over scarce aircraft and maintenance resources. These data elements included the job start and job stop times and the number of men involved in a maintenance task. Additional important information included the reason for jobs being delayed, either in starting or in continuing to completion.

It took a number of years and much demonstration, especially through field tests, to secure the inclusion of these data elements into the USAF Maintenance Management System (as contained in Air Force Manual AFM 66-1). Aside from the need to demonstrate improvement in management, there is always the difficulty present of introducing change in large organizations, with the need to secure agreement among all commands, to change the instruction manuals, and to print new forms in the millions. All of this effort takes time.

With these additional data elements as part of the AFM 66-1 system, it is possible to produce time traces that show the minute by minute maintenance status of an aircraft at base level. The focus of the maintenance manager is on the maintenance status, and the time trace can show if an aircraft is in maintenance, with no maintenance being performed on the aircraft. This kind of display is important because the presumption is that the goal of a maintenance organization is to complete maintenance on a multimillion-dollar aircraft as rapidly and and as efficiently as possible, and management should be concerned if the display shows that substantial periods of time occur with no maintenance being performed. This type of display can be used to show the length of delay as well as the reasons for the delay. At this point, the full benefits of these displays are not being realized because they do rely on computer programs that have not been made a standard requirement for all Air Force organizations. However, they have been used in many test situations.

Such display information tends to be diagnostic in nature. Over time, it helps to identify the amount of delay and the causes of delay. It helps management to make changes in resource allocation, work schedules, inventory policy, and so on, that can help to reduce the delays, and thereby presumably speed up the turnaround time for aircraft passing through maintenance. Various scheduling rules and goal-oriented techniques

have been developed for use along with these displays, but these methods have been primarily exploited in research and test situations.

The diagnostic usefulness of such display information needs to be supplemented by other decision-making tools, particularly models that provide guides to longer-term resource allocation. Such planning models are especially useful in dealing with steady state or essentially equilibrium conditions. They recognize that there must be a tradeoff between maintenance responsiveness and the resources assigned to maintenance on a long-term basis. Models of this nature can be quite aggregate or they can be highly detailed.

Both SAMSOM and L-COM, two major modelling efforts done at Rand and the Air Force in the mid-1960s, are examples of detailed simulation models. These models deal with the tradeoffs between operational capability and logistics resources. They made demands on the standard data systems for detail that was not then available. Thus, it was early experience with SAMSOM during the mid-1960s that stressed the need for the augmented AFM 66-1 data system, particularly on team-size detail, for each maintenance job. And L-COM, which is a network model, required construction of the aircraft maintenance task networks down to the black box or module level, and to specify for each task the specific maintenance AFSCs (Air Force Specialty Codes) used. Both these models, particularly L-COM, have been used in the Air Force, and for purposes beyond those originally intended. It is understood that L-COM has been validated against real world experience with sufficiently satisfactory results that it is now being used to develop resource requirements for maintenance personnel, and ground support equipment, under a variety of operational conditions for mature weapon systems, as well as for some newer weapon systems.

Thus, models have also helped establish additional data collection needs, because such data are necessary to permit the use of such models in planning and other functions of management. Realization of the full benefits from these additional data elements in day-to-day base-level operating management in performing such activities as maintenance scheduling, job dispatch, manpower control, and so on, will have to await further

progress in the availability of computer-based maintenance management and control systems.

5.4 Scheduling Example

A significant tool for resource allocation is known as scheduling. By scheduling we mean the assignment of resources over time in such a way that stipulated goals of the organization are being satisfied. Scheduling can be done for varying time periods. Some schedules are prepared to cover a month's activities, giving day-by-day detail. Others are prepared for each day to show the specific actions and assignments for the day.

One key type of scheduling problem important to logistics is that concerned with the operation of aircraft at airbase level. Within a command, such as the Strategic Air Command, bomber aircraft are flown for aircrew proficiency training. The scheduling work involves two major components: (1) the so-called operations scheduling which involves the specification of the training to be done on each sortie, taking into account the status of training of each of the crews; and (2) maintenance scheduling, which involves the selection of aircraft for each training mission and the scheduling of the maintenance tasks and resources for making the aircraft mission-ready.

This discussion is oriented toward this kind of scheduling problem because it contains most of the complex elements that must be treated in making progress on scheduling, especially in using the computer, because the current capability, at least within the Air Force, is dependent on manual forms of scheduling. Such limited computational capability restricts the options open to the scheduler in seeking to improve the quality or performance of his schedules.

The scheduling theory and its implementation for operations-maintenance scheduling are in a relatively early and developmental state. This work is truly a research and development problem of the 1970s, with antecedents in the late 1960s. Its complexity arises because a schedule must satisfy both objectives and constraints, and these are sometimes mutually contradictory or at least interacting in relatively subtle ways. Scheduling of aircraft operations for training purposes, for example, has historically been a heuristic

process, with much of the work done manually. The result is that scheduling is recognized as a tortuous process that permits very little examination of alternatives in terms of objectives or allocation of resources.

Current aspirations for advancing the state of scheduling capability are to provide a man-machine environment, in which the scheduler can experiment with alternative schedules, trying to fit in various circumstances that could affect either the aircraft status or the people involved in operations and maintenance. Since such a nonspecific condition characterizes scheduling, it is necessary to provide flexibility to the scheduler and to permit him to give more or less weight to different parts of the scheduling interactions. This manipulation of the schedule is a way of making goal-setting decisions and observing the consequences on the schedule. This weighting of different aspects of the schedule basically relates to notions of utility as perceived by the scheduler. The scheduler tries to bring in his utility preferences to reflect specified objectives or requirements, but because of constraints inherent to the performance of the scheduled activities, the heuristic process must take over to resolve the interactions that result. However, the resulting schedule may or may not please him, so the scheduler needs a capability to change it.

The scheduling process is thus a continuous one, since conditions and even goals themselves might change, so that the scheduler is always assessing the situation against these goals, seeking to make progress in efficient and consistent ways. This effort is aided by various display techniques and analyses reflecting status and problems.

This type of scheduling process, with its support of choices, needs a rich data base. In the case of combat crew training, it requires much detail on the status and characteristics of crew members, the composition of training curricula, the outcomes of training exercises and tests, and much information on each training aircraft as well as the resources used to maintain the aircraft.

In the initial formulation of the scheduling process, it had been thought that the development of the

underlying information system could be undertaken independently of the development of the scheduling algorithms. This assumption was made because it was felt that the existence of a data base that could be queried on-line would be a useful tool for the managers, and it could also be done relatively early in the development process. However, the interlocking nature of the queries and the scheduling process has become so evident that the old concept has been dropped, and it is now accepted that there must be the concurrent development of the scheduling algorithms and the associated data system.

Scheduling has significant implications of data presentation and analysis for a variety of reasons. For one thing, it is viewed as an interactive process between man and the computer. This means that access to the computer should be made relatively simple and easy since the schedulers will not typically be computer specialists. Second, output from the computer in terms of displays should be easy to read and be helpful in providing ways to improve schedules. The art and technique of displays in the man-machine process deserves special recognition in successful implementation of scheduling systems.

Basically, what it would be nice to have is a scheduling black box into which the scheduler puts his objectives or utility preferences related to the tasks to be scheduled, and out of which come schedules that he can modify rapidly as he tries to resolve the conflicts in goals and constraints. The black box would contain all the current data, heuristics, special constraints, and so on, needed by the scheduler to obtain feasible schedules, and his efforts are then focused on improving the schedule's quality. This kind of system is not easy to achieve, but if it could be done, the scheduling capability will have advanced a great deal.

"What if" questions should also be made easy in an interactive scheduling system. The purpose of a schedule is to accomplish certain objectives, and so a schedule, if followed, can show outcomes that serve as responses to "what if" questions.

As is clear, this scheduling theory and study effort, along with its information system complement, is an evolving and not particularly advanced research area.

This means that much of the early effort must be undertaken by means of experimental and evolutionary prototype systems that will help to suggest further steps in progress to truly implementable systems that have the capabilities required in an operating situation. From exposure to various studies on scheduling and other types of logistics management systems involving man-machine interactions or other complex organizational decision-making activities, the importance of using prototypes to learn about the desirable and undesirable characteristics of such systems cannot be overstressed.

5.5 Data Acquisition

The history of logistics data has been strongly focused on the process of data acquisition. The creation of systems in the Air Force such as AFM 66-1 and AFRAMS (the Air Force recoverable asset management system) has been marked by management efforts to ensure complete reporting, accurate data, and hopefully, effective use of the data products.

The advent of built-in automatic reading and recording systems (usually referred to as AIDS, standing for "airborne integrated data systems") has made it much easier to collect data on aircraft conditions in flight. One intent in using AIDS is to reduce aircraft downtime by more rapid reporting and diagnosis. As a matter of fact, with the introduction of AIDS, the situation may be becoming one of data congestion, which could add significantly to the costs of these systems because of the heavy requirements for data processing. With so much data to analyze, it is understood that difficulties are being encountered in using these data rapidly for diagnostic purposes. Such techniques as trend analysis do not yet seem to have the predictive properties that had been anticipated, except in limited instances. Furthermore, it appears as if the problem of using these data for diagnostic purposes may call for basic kinds of engineering knowledge that may not yet be fully in hand.

It seems as if the problem is one of developing greater understanding about the basic failure processes so that a model of the physical environment can be used to generate the data requirements and to interpret the diagnostic data in ways that improve logistics support. There has been some experience on the problems of data

collection oriented toward these engineering and logistics studies through special field tests on specific aircraft types and in particular operational environments. For the time being, it seems that these diagnostic efforts must be tailored to specific problem situations because of the lack of basic technological understanding about the failure process. Under this learning situation, it is difficult to standardize the analysis process for which automatic systems are typically designed.

The point of this discussion is to raise some questions about the use of expensive and also possibly unreliable sensor or other recording equipment to collect large quantities of data when more understanding is required to make effective use of the data. It would seem that there is a need to invest more effort into research on failure analysis oriented to explaining the failure process in physical and engineering terms, so that the data and interpretation produced by such recording and testing equipment are valid and reliable, with resulting important logistics benefits.

The logistics implications of what is being said about such automatic recording systems can be illustrated by the following. With the current use of integrated systems, especially in avionics, incorrect diagnosis can cause the removal of black boxes, which when returned to some repair facility are found to function perfectly upon being checked. This means that the repair pipeline contains serviceable modules that tend to inflate demand statistics and pipelines, with consequent increased procurements of extra spares, as well as additional expenditure of man-hours, transportation, and so on.

Limited experience from previously mentioned field tests with integrated avionic systems indicates that such RTOK (returned, tested OK) modules may occur not only because of hardware deficiencies but because of limitations in the software used in AIDS to diagnose and interpret the condition of the individual but integrated hardware elements. It is important to understand whether this inadequacy could reflect an unsatisfactory "model of the world" as contained in the software rules, or whether the diagnostic routines which are also in the software need improvement. Experience

with such problems is too limited to provide definitive answers to a situation that involves a combination of analytic, engineering, and logistics expertise. These multidisciplinary problems are becoming more central to analysis, control, and operation of logistics systems.

5.6 Exploitation of Data Variability

This technique, which is an ancient approach to data analysis, has been exploited in important ways in logistics. The basic notion in science is that of experimental design. One establishes a design in which the treatments accorded different subjects or objects are varied in a pre-established way to help evaluate the effect of the treatment on the outcomes being measured. Now for many laboratory and other controlled situations, this scientific approach can be followed, and it has been very successful.

In institutional or real world situations, such as logistics, where there are complex phenomena in which the interactions are very imperfectly understood, so that it is difficult to establish a so-called controlled experiment, opportunities can be found to exploit the types of variability that occur naturally through the replication of particular events. This approach can be especially useful when one is dealing with problems in which the theory for explaining causes is lacking or not understood. This situation is not infrequent in the logistician's world. In the absence of theory, decisions on physical activities are sometimes made through policy declarations that in effect are subjective determinations of what seems best to do under the circumstances. Unprogrammed or inadvertent deviations from such policy, which could be the result of many factors, provide real world experience that may offer opportunities for evaluating the established policy.

A good example of this experience can be provided from logistics. The time interval between major inspections of an aircraft at a depot is based on policy that contains subjective elements. Changes in the inspection interval are made from time to time, presumably relying on experience obtained in the course of maintenance, as well as the number of aircraft that are in the depot, and therefore the resources that must be provided to this activity. As the interval is extended, fewer

aircraft are in the depot, and the maintenance capacity required is reduced. This depot inspection and repair activity is not inconsequential from a resource viewpoint. Therefore, there are good reasons to seek judicious extension of the interval.

As has been said, interval setting and extension is somewhat a subjective matter, and an attitude of conservatism seems to be followed in changing such intervals. For one thing, there is a desire to retain resources, since this provides a hedge against sudden increases in workload. In addition, there is a sense that it is safer to have aircraft receive depot inspections more often rather than less often. On the other hand, there are worthwhile reasons to extend intervals more rapidly if evidence suggests it should be done.

Turning now to the theme of data variability using the aircraft interval problem for illustration, it had been first thought that it would be necessary to conduct special experiments to get information on possible interval extensions. This would have been expensive and even difficult to do technically, aside from convincing the institution that such experimentation is desirable and useful. However, when empirical data were studied, it was found that through various circumstances there were aircraft that had exceeded the policy interval by varying amounts, some substantially. Thus, by this "accident" the analyst had access to aircraft that had experienced longer intervals than dictated by policy. The analysis of these aircraft, in terms of work to be done, condition, and so on, showed no significant difference in their requirements from those that had received their inspections at the specified interval. Consequently, in an empirical manner, there was evidence that it would be appropriate to consider an interval extension.

Obviously, in addition to exploiting existing variability, there is the further option of inducing interval variability, so that experience can be obtained at inspection intervals beyond those specified in current policy. This approach involves experimental design, using induced variability to establish the desirability of extending the inspection interval and the amount of extension. This subject is an extremely complicated one because there is no evidence pointing to one "best"

interval, at least from an engineering standpoint. Even the approach taken by the airlines is also an empirical and subjective one. As one looks at airline data and analyzes the way in which intervals have been extended on successive generations of jet aircraft, starting with the Boeing 707, then the 727, and now the 747, one sees a more rapid extension of intervals for comparable stages of life with each of these aircraft.

The study of interval extension opens up other aspects of analysis. When one examines a population of aircraft that is subjected to interval extension, there is a so-called "maturation phenomenon" involved. Thus, if the inspection interval is extended from two to three years, it takes a full six years before all aircraft are on the new inspection interval. Therefore, a decision to move to an extended interval does not subject the entire aircraft fleet immediately to whatever unforeseen dangers there may be in the longer inspection interval. However, there is the immediate benefit from the reduction in resources required, in this case a reduction of about one-third.

It is possible, if trouble develops, to stop the interval extension process, and reduce the interval. Of course, just as it takes time to bring all aircraft up to the new interval, it also takes time to return all aircraft to the old interval. From what has already been said, there is evidence to indicate that interval extension policies tend to be conservative, so that the likelihood of needing to set them back, in the current environment, is small.

This subject of interval extension provides an interesting illustration of empirical data analysis, combined with some relatively modest mathematical modelling of simple dynamic relationships as affected by changing aircraft inspection intervals. Use of this model shows that it takes a long time for the effect of an interval change to increase the inspection interval for the whole fleet of aircraft that go to the depot for periodic inspection. It would be desirable to consider, or at least contemplate, a more dramatic approach to periodic inspection in order to realize more rapidly the benefits of interval extension. One possible way to achieve such a result is to view the

depot inspection problem as one of sampling aircraft of different ages, measured in terms of the interval since last inspection. This approach would permit the introduction of interval variability in a systematic way, and it may avoid the lengthy maturation phenomenon that occurs when interval extension is treated as a gradual fleet-wide policy.

Thus, one possibility is that the problem can be approached from a sampling context, recognizing that the appropriate inspection interval is an uncertain or unknown policy parameter. A strategy could be developed in which a relatively small proportion of the aircraft are sent in for inspection at selected intervals, and an analysis made of their condition. Both the proportion inspected and the interval between samplings could be varied depending on the information obtained and the sampling and inspection strategy followed. It is presumed that the sample size selected would be reasonably small, say, on the order of 10 percent, so that if the samples were done annually, this would be equivalent in numerical terms to a 10-year interval.

Such a sampling approach would drastically change the approach taken to scheduled maintenance. Instead of treating depot visits by aircraft as prescheduled opportunities to do work, the particular aircraft would probably be selected randomly from a sample of aircraft that were of a given age or interval since last depot inspection. Such a technique is now being used to obtain engineering data on the condition of aircraft, but it is not tied in an explicit way to the inspection policy. This proposal would combine the sampling policy and the inspection policy into a single process that would affect the scheduled maintenance workload at base and depot, depending on the sampling outcome.

Obviously, such an opportunistic and adaptive approach to depot inspection does call for flexibility in the depot's capacity to absorb changing workload demands, but this is the usual environment in which the depot operates. Time after time, sudden workload demands occur, be they major modifications or special projects dictated by operational requirements. Therefore, if the approach is found to be useful for managing scheduled maintenance, the depot should be able to handle the consequence of some unpredictability in this type

of workload.

This illustration is given to show the opportunities for imaginative approaches to old problems. There are other interesting parts to the scheduled maintenance problem, including its relationship to unscheduled maintenance, to modification activity, and to the location of maintenance activities, all of which are deserving of detailed investigation since they could have large impacts on readiness, manpower savings, and changes to the logistics structure.

5.7 A Perspective

Thus far, this survey has dealt with specific examples of how data affect logistics decision-making. One is always tempted to generalize from these specifics to try to obtain an overall perspective on where this effort is tending, hoping that more structured insights will provide a productive opportunity for progress.

As the examples have indicated, initially much of the focus was on the use of data for analysis, largely statistical, to help identify the key variables of relationships affecting logistics requirements and performance. Such analysis led to the creation of more formal and elaborate models of the phenomena. This work was aided by such concepts as tradeoffs and cost-benefit analysis, as well as viewing logistics as a system. Such views led to the development of models for studying these tradeoffs. Thus, in the inventory area, there has been much interest in the tradeoff between stock levels and resupply time, or tradeoffs among supply, maintenance and transportation of spare parts. In maintenance, among the tradeoffs studied has been that between percent of aircraft NORM (not operationally ready maintenance) and the amount of maintenance resources available, taking into account the aircraft activity rates and perhaps other factors.

Such models were actively built and used in the late 1950s and 1960s. Their construction and use were aided by the developing capabilities of computers and various programming languages. These models undoubtedly helped to shape some of the developments with logistics systems and to require the use of more analytically oriented approaches to logistics decision-making. Their use also led to a requirement for many military

personnel trained in the application of these techniques, so that models are now becoming a more substantial part of the study and analysis process used for internal decision-making.

Data analysis helped to shape the contents of models, and in turn the models helped to determine data requirements that were not already in the ongoing data collection systems. Such data requirements were sometimes embedded in the routine data collection systems, or at times special field tests or experiments were employed to obtain data required for the models.

Most cost-effectiveness models are largely used for planning purposes, that is, they help in those decisions or policies that lead to relatively long-term allocations of resources. The models are also used for sensitivity testing purposes, since the effects of such long-term decisions should be relatively stable over the range of conditions likely to be encountered in the future.

The next stage in the evolution of the use of information for decision-making, and one that is more recent in its general interest, is that associated with management control. These uses attempt to make decisions based on current information, and the impact of these decisions tends to be shorter-run in effect. Management control systems involve much interaction between the information system and the decision-maker on a frequent basis, because the problems encountered are dynamic, requiring the decision-maker to make constant adjustments to his plan. The scheduling example discussed previously is typical of this kind of problem.

The state of the art of management control systems is probably less developed than the traditional analysis and modelling previously discussed. The former subject beings up difficult elements of decision-making, such as utility measures and time discounting, that is, the degree to which the decision-maker is prepared to discount the future in making decisions about the present. These subjective determinations cannot be automated, but rather it is necessary to permit the decision-maker to introduce his judgment about such factors dynamically into the decision process.

This aid to the decision-maker must be provided not only through the specific items of information supplied

to him, but also through the way in which they are portrayed for his consideration. Information displays, formats of representation, and various triggering or flag devices for alerting the decision-maker to changed circumstances are forms of data representation that management control systems now require. The techniques for devising these formats and evaluating them are now required as part of this new technology.

These man-machine techniques are in early stages of development because they do involve the interactive activity of man and computer in decision-making activity. The principles of this capability involve heuristic processes that must be created as part of the control system design process.

In order for the control system to be useful to the decision-maker, it is necessary that he have a way to inject his utility function into the control system, and then observe the results of his actions. Thus the control system itself becomes a set of heuristics that operate on the decision-maker's utility choices, and then produce resulting decisions and actions. Since the decision-maker is usually not explicit about his objectives, or the problems are complex, he needs a flexible and responsive control system. Without such capability, the decision-maker's motivation or ability to use such control systems is not high because they do not seem to fit his needs.

There has been some progress through laboratory studies and field tests on understanding how the decision-maker in logistics can be motivated to use the computer in his control process. It has also been learned through experience that it is difficult and costly to study this problem with full-blown control systems. Failures with big control systems are legion in the management information and control system field. The limited successful experience with this type of problem suggests the advantageous use of prototypes, which provide a means for system designers to learn how the computer can be most helpful to the decision-maker by trying to involve him specifically in the design process in comparatively realistic ways. Thus, the man-machine interface is itself a parameter of system design which can vary in different parts of the system and at different phases of the design process, depending on the

ability of the designer to meet the decision-maker's needs. Such insights into the man-machine aspects are required so that the computer's role can be defined in sufficient detail to provide the software, information inputs and outputs, and other design characteristics of a control system useful to a decision-maker. This statement contains an assessment of many years' observing and working with logistics management systems. It tries to be helpful and systematic about what can be accomplished, but past realities still indicate difficult experience lies ahead.

From the ranging nature of this survey, it can be seen that the subject has many facets. In the final part, an effort has been made to pull the pieces somewhat together to indicate how the processes of data collection, representation, and analysis interact in logistics decision-making. Although much progress and understanding have been achieved, the problems and challenges are still tremendous. The need for research, systematic study, and appraisal of what has been accomplished and learned will continue. The payoffs in logistics will be both increased support capability and more efficient use of resources.

Selected Rand Logistics Bibliography

Astrachan, M., and C. C. Sherbrooke (1964). An empirical test of exponential smoothing. The Rand Corporation, RM-3938-PR, (March).

Bell, C. F., and T. C. Smith (1962). The Oxnard base maintenance management improvement program. The Rand Corporation, RM-3370-PR, (November).

Berman, M. B. (1974). Improving SAC aircrew and aircraft scheduling to increase resource effectiveness. The Rand Corporation, R-1435-PR, (July).

Brooks, R. B. S., C. A. Gillen, and J. Y. Lu (1969). Alternative measures of supply performance: fills, backorders, operational rate, and NORS. The Rand Corporation, RM-6094-PR, (August).

Brown, B. B. (1956). Characteristics of demand for aircraft spare parts. The Rand Corporation, R-292, (July).

Campbell, H. S. (1963). Concept and measurement of demand for recoverable components. The Rand Corporation, RM-3824, (September).

Cohen, I. K. (1972). Aircraft planned inspection policies: a briefing. The Rand Corporation, R-1025-PR, (June).

Cohen, I. K., E. V. Denardo, and P. J. Kiviat (1966). Integrating base maintenance management by unifying its information systems in manual and computer-assisted environments. The Rand Corporation, RM-4849-PR, (June).

Cohen, I. K., O. M. Hixon, and B. G. Marks (1966). Maintenance data collection and workload control information systems: a case study. The Rand Corporation, RM-4985-PR, (November).

Cohen, I. K., O. M. Hixon, and R. L. Van Horn (1965). Unifying resource allocation, control, and data generation: an approach to improved base-level maintenance management. The Rand Corporation, RM-4778-PR, (November).

Conway, R. W. (1964). An experimental investigation of priority assignment in a job shop. The Rand Corporation, RM-3789-PR, (February).

Dade, M. A. (1973). Examples of aircraft scheduled maintenance analysis problems. The Rand Corporation, R-1299-PR, (December).

Drezner, S. M., and R. L. Van Horn (1967). Design considerations for CAMCOS--a computer-assisted maintenance planning and control system. The Rand Corporation, RM-5255-PR, (July).

Fisher, R. R., W. W. Drake, J. J. Delfausse, A. J. Clark, and A. L. Buchanan (1968). The logistics composite model: an overall view. The Rand Corporation, RM-5544-PR, (May).

Geisler, M. A., B. B. Brown, and O. M. Hixon (1954).

Analysis of B-47 consumption data and activity. The Rand Corporation, RM-1288, (July).

Geisler, M. A., W. W. Haythorn, and W. A. Steger (1962). Simulation and the logistics systems laboratory. The Rand Corporation, RM-3281-PR, (September).

Karr, H. W., M. A. Geisler, and B. B. Brown (1955). A preferred method for designing a flyaway kit. The Rand Corporation, RM-1490, (May).

McIver, D. W., A. I. Robinson, H. L. Shulman, and W. H. Ware (1974). A proposed strategy for the acquisition of avionics equipment. The Rand Corporation, R-1499-PR, (December).

Miller, L. W. (1972). A simple adaptive scheduling mechanism for planning base level inspections. The Rand Corporation, R-938-PR, (February).

Miller, L. W. (1973). VIMCOS II: a workload control simulation model for exploring man-machine roles in decisionmaking. The Rand Corporation, R-1094-PR, (June).

Miller, L. W., A. S. Ginsberg, and W. L. Maxwell (1974). An experimental investigation of priority dispatching in aircraft maintenance, using a simplified model. The Rand Corporation, R-1385-PR, (June).

Miller, L. W., R. J. Kaplan, and W. Edwards (1968). JUDGE: a laboratory evaluation. The Rand Corporation, RM-5547-PR, (March).

Sherbrooke, C. C. (1968). A management perspective on METRIC--multiechelon technique for recoverable item control. The Rand Corporation, RM-5078/1-PR, (January).

Smith, T. C. (1964). SAMSOM: support-availability for multi-system operations model. The Rand Corporation, RM-4077-PR, (June).

Sweetland, A., and F. Finnegan (1969). The RAND/TAC information and analysis system: volume I--data

collecting and editing. The Rand Corporation, RM-5666-PR, (January).

Van Horn, R. L., A. S. Ginsberg, and K. Merrill, eds. (1965). Production scheduling and control: proceedings of a joint Air Force/Rand symposium. The Rand Corporation, RM-4576-PR, (July).

Youngs, J. W. T., M. A. Geisler, and B. B. Brown (1955). The prediction of demand for aircraft spare parts using the method of conditional probabilities. The Rand Corporation, RM-1413, (January).

Chapter 6

THE DESIGN OF LARGE SCALE LOGISTICS SYSTEMS: A SURVEY AND AN APPROACH*

Arnoldo C. Hax
Massachusetts Institute of Technology

6.1 Introduction

Much has been said about the lack of impact management science has had in providing effective support to managers, particularly at the top levels of the organization. To some degree, this is due to the insufficient knowledge on the part of managers who are unable to understand the strengths and limitations of mathematical models, and to the naivete of management scientists who fail to appreciate the complexities of the management process.

Large systems, which require sophisticated computer assistance to operate, are usually expensive and time consuming to design. If they are going to serve a useful purpose, their development demands a great degree of interaction between the managers they are intending to help, and the management scientists and computer experts who are responsible for their technical design. In practice, this interaction is very hard to accomplish and most systems fail due to the inability of the parties involved to communicate effectively with one another. Moreover, when the system is to support operational decisions, even if the initial design is done properly, system maintenance, expansion, and updating still remain as great problems. Due to the fast rate of change of personnel, so typical in today's business firms, very often those originally involved in the system design are quickly moved up or out of their previous position leaving behind a complicated apparatus that is hard for their successors to understand and manipulate. This accounts for many cases of initial success and eventual failure.

*The preparation of this chapter was supported in part by the Office of Naval Research under Contract N00014-67-A-0204-0076 with the Massachusetts Institute of Technology.

Some of these problems can be resolved by a customized approach to system design, where the user is confronted with a questionnaire that allows for a quick representation of the user's understanding of his decision support requirements and environment, and where the system responds with a diagnostic of the user's needs and with a suggestion for a specific system that is responsive to the user's requirements. This process facilitates the user's education in the system capabilities, and constitutes an appropriate way to introduce modifications resulting from changes in the decision environment.

The present chapter attempts to formulate an approach to the design of a large-scale, customized, model-based system to support logistics decisions. The field of logistics has produced significant contributions in the use of models to enhance managerial decision-making capabilities, particularly in the areas of procurement, production planning, inventory control, and distribution. We believe the field is ripe for the development of this comprehensive approach.

We begin by analyzing the logistics decision process. In Section 6.2 three different taxonomies are presented to classify logistics decision . Each classification leads to some important conclusions with regard to the characteristics of the logistic support system. Section 6.3 provides a review and a critique of existing computer based logistics systems. Special attention is given to the work of Connors et al. [6.16] and the IBM System/3 Customizer [6.53]. Both of these approaches constitute important steps in the direction of system design advocated in this paper. Finally, Section 6.4 lists the essential characteristics we feel a logistics decision support system should possess and suggests a way to facilitate the implementation of these concepts.

Throughout this chapter we are emphasizing the design of systems to support day-to-day ongoing logistics decisions, as opposed to major decisions which occur only occasionally in the life of an enterprise. Moreover, most of our remarks are directed toward a traditional manufacturing firm operating in the private sector. However, we believe our conclusions can be extended easily to support logistics decisions for the public sector.

6.2 The Logistics Decision Process

Logistics is concerned with the effective management of the total flow of goods, from the acquisition of raw materials to the delivery of finished products to the final customer. A logistics system is composed of a large number of elements which have to be managed effectively in order to deliver the final products in appropriate quantities, where they are required, at the desired time and quality, and at a reasonable cost. In a general situation the most important elements are the following.

Multiple plants, possibly containing a wide variety of equipment that represents the manufacturing capabilities of the firm

Multiple warehouses (distribution centers, local, regional, and factory warehouses) that might define a complicated distribution network

Multiple products (including raw materials, supplies, semifinished and finished products) to be purchased, manufactured, and distributed by the organization (The product structure could be fairly complex involving several production stages of fabricating and assembly operations. The total number of individual items, representing varieties of product specification, color, dimensions, and so on, could exceed several thousand.)

Transportation and local delivery means, either owned, leased, or contracted, by the firm

Communication and data processing equipment

People, representing a wide variety of skills covering all the organizational echelons

The acquisition and utilization of these elements are subject to a wide variety of constraints. Examples are the following.

Manufacturing and distribution characteristics

Productivity constraints

Equipment capacity constraints

Labor availability

Technological constraints

Purchasing, manufacturing, and distribution lead times

Demand uncertainties and seasonalities

Service requirements

Other constraints (institutional, financial, marketing, and so on)

In order to determine effective ways to acquire and use the logistics elements subject to these constraints, several cost components have to be taken into consideration. The most important factors contributing to cost are the following.

Capital investments

Production and purchasing costs

Setup or changeover costs

Purchase ordering costs

Transportation and handling costs

Hiring and firing costs

Inventory related costs

Promotional and advertising costs

From the mere enumeration of these logistics system components, it is clear that the underlying decision process can be extremely difficult. Decisions involve comparisons of a large number of alternatives with

complex interactions among system components. They cover several organization echelons forcing a great deal of coordination both vertically and among functional areas. The exclusive use of conventional managerial wisdom, based largely on experience and intuition, may not be adequate in assuring effective decision-making. We will explore ways in which to support the management actions by means of model base systems.

To identify the characteristics that a sound support system should have, we will start by reviewing three different frameworks that have been proposed to classify logistics decisions. These frameworks provide simple taxonomies, stressing the relative differences of some aspects of the logistics process. At times, they tend to oversimplify and overgeneralize the inherent complexities of the process. However, we feel important inferences can be drawn from these frameworks which we are going to exploit in setting up design criteria for logistics support systems.

6.2.1 Anthony's Framework: Strategic, Tactical and Operational Decisions. The first of these frameworks was proposed by Robert N. Anthony [6.1]. He classified decisions into three categories: strategic planning, tactical planning, and operations control. Let us briefly comment on the characteristics of each of these categories and review their implications for a model-based approach to support logistics management decisions.
Strategic Planning: Facilities Design. Strategic planning is concerned mainly with the establishment of managerial policies and with the development of the necessary resources the enterprise needs to satisfy its external requirements in a manner consistent with its specific goals. In the area of logistics we are considering, the most important strategic decisions are concerned with the design of the production and distribution facilities, involving major capital investments for the development of new capacity, either through the expansion of existing capacity or the construction or purchase of new facilities and equipment. These decisions include the determination of location and size of new plants and warehouses, the acquisition of new production equipment, the design of working centers within

each plant, and the design of transportation facilities, communication equipment, data processing means, and so on. Other decisions of this nature which have significant marketing and financial implications are make-or-buy decisions, product line diversity, quantity versus price tradeoffs, and divestment of facilities.

These decisions are extremely important because, to a great extent, they are responsible for maintaining the competitive capabilities of the firm, determining its rate of growth, and, eventually, defining its success or failure. An essential characteristic of these strategic decisions is that they have long-lasting effects, thus forcing long planning horizons in their analysis. This, in turn, requires the consideration of uncertainties and risk attitudes in the decision-making process.

Moreover, investments in new facilities and expansions of existing capacities are resolved at fairly high managerial levels, and are affected by information which is both external and internal to the firm. Thus, any form of rational analysis of these decisions has of necessity a very broad scope, requiring information to be processed in a very aggregate form to allow all the dimensions of the problem to be included and to prevent top managers from being distracted by unnecessary operational details.

Tactical Planning: Aggregate Capacity Planning. Once the physical facilities have been decided upon, the basic problem to be resolved is the effective allocation of resources (for example, production, storage and distribution capacities, work force availabilities, and financial and managerial resources) to satisfy demand within technological requirements, taking into account the costs and revenues associated with the operation of the production and distribution process. When dealing with several plants, with many distribution centers, regional and local warehouses, with products requiring complex multistate fabrication and assembly processes, affected by strong randomness and seasonalities in their demand patterns, these decisions are far from simple. They usually involve the consideration of a medium-range time horizon, divided into several periods, and the aggregation of the production items into product families. Typical decisions to be made within this

context are utilization of regular and overtime work force, allocation of aggregate capacity resources to product families, accumulation of seasonal inventories, definition of distribution channels, and selection of transportation and transshipment alternatives.
Operations Control: Detailed Production Scheduling. After making an aggregate allocation of capacity among product families, it is necessary to deal with the day-to-day operational and scheduling decisions which require the complete disaggregation of the information generated at higher levels into the details consistent with the managerial procedures followed in daily activities. Typical decisions at this level are: the assignment of customer orders to individual machines; the sequencing of these orders in the work shop; inventory accounting and inventory control activities; dispatching, expediting, and processing of orders; and vehicular scheduling.

These three types of decisions differ markedly in various dimensions. The nature of these differences expressed in relative terms, is characterized in Table 6.1.

6.2.2 Implications of Anthony's Framework: A Hierarchical Integrative Approach. These are significant conclusions that can be drawn from Anthony's classification regarding the nature of the model-based decision support systems. First, strategic, tactical, and operational decisions cannot be made in isolation because they interact strongly among one another; therefore, an integrated approach is required if one wants to avoid the problems of suboptimization. Second, this approach, although essential, cannot be made without decomposing the elements of the problem in some way, within the context of a hierarchical system that links higher level decisions with lower level ones in an effective manner, and in which decisions that are made at higher levels provide constraints for lower level decision-making. This hierarchical approach recognizes the distinct characteristics of the type of management participation, the scope of the decision, the level of aggregation of the required information, and the time framework in which the decision is to be made. In our opinion, it would be a serious mistake to attempt to

Table 6.1 Summary of Anthony's Framework

	Strategic Planning	Tactical Planning	Operations Control
Objective	Resource acquisition	Resource utilization	Execution
Time horizon	Long	Middle	Short
Level of management involvement	Top	Medium	Low
Scope	Broad	Medium	Narrow
Source of information	(External and	Internal)	Internal
Level of detail of information	Highly aggregated	Moderately aggregated	Detailed
Degree of uncertainty	High	Medium	Low
Degree of risk	High	Moderate	Low

deal with all these decisions simultaneously, via a monolithic system or model. Even if computer and methodological capabilities would permit the solution of a large detailed integrated logistics model, which is clearly not the case today, that approach is inappropriate because it is not responsive to the management needs at each level of the organization, and would prevent the interactions between models and managers at each organization echelon.

In designing a system to support management decisions, it is imperative, therefore, to identify ways in which the decision process can be partitioned, to select adequate models to deal with the individual decisions at

each hierarchical level, to design linking mechanisms for the transferring of the higher level results to the lower hierarchical levels that includes ways to disaggregate information, and to provide quantitative measures to evaluate the resulting deviations from optimal performance at each level.

In Section 6.4 we suggest some criteria to accomplish the hierarchical system design.

6.2.3 A Traditional Framework: The Logistics Managerial Functions. Our previous framework was useful to relate the logistics decision process to the organizational structure of the firm and to stress the need to design a decision support system that is both integrated and hierarchical. We also saw the large number of individual functions that take place in managing the wide variety of elements present in the logistics activities.

In fact, the most traditional approach used to classify logistics decisions is based on a managerial function taxonomy. The majority of production and logistics books have their chapters organized according to these managerial functions. Since a significant amount of research has been devoted to studying specific problems pertaining to these functional areas, it is useful for us to review briefly this classification. The most important logistics managerial functions are the following.

Forecasting generates long, middle, and short-term projections of expected demand for each item (or families of items) that is purchased, produced, and distributed by the firm.

Facilities design involves the determination of number, location, and size of plants and warehouses to supply the market requirements.

Aggregate production planning deals with the allocation of capacity, work force, and inventories to satisfy aggregate demand through a middle-range planning horizon.

Inventory control decides on how much and when to

order each purchased or manufactured item in accordance with the service requirements specified by the firm and the constraints imposed by aggregate production planning decisions.

Production scheduling involves the assignment of men and machines to specific operations during short time intervals (usually daily schedules), subject to the higher level decisions resulting from aggregate production planning and inventory control.

Purchasing determines when and how much to order of each externally supplied item. The procurement of raw material might also be constrained by aggregate production planning decisions.

Distribution concerns the availability of products throughout the production and distribution network, transferring inventories from one location to another so as to best satisfy the demand requirements at each location.

Vehicular scheduling deals with the allocation and routing of vehicles to provide appropriate delivery of goods to the prescribed production and distribution locations.

Information processing organizes the transmission of information throughout the organization echelons. The basic input to the logistics information processing function is the customer order. Thus, of central importance for this function is the development of an appropriate customer order processing cycle, which includes the acknowledgment, validation, editing, processing, invoicing, and shipment of the order.

We have purposely omitted several other functions from this enumeration that are closely related to logistics but fall more properly into the category of industrial engineering activities--for example, maintenance, quality control and quality assurance, equipment reliability and replacement, plant layout, work improvement, and standardization. It is also important to emphasize the need of coordinating the logistics functions with the

remaining managerial functions of the firm, such as personnel, marketing, finance, engineering, and sales.

6.2.4 Implications of the Managerial Functional Framework: A Bag of Tricks. As we have indicated above, most of the traditional work in the application of management science to logistics has been oriented toward the solution of well-defined problems belonging to the functional areas just described. Part of the skill involved in designing an integrated logistics support system is to partition the overall problem into subproblems that can be effectively solved within the current state of the art. The management functional framework can be very helpful to identify meaningful subproblems and to associate with these subproblems the existing tools that are available for their solution. The resulting catalog of problems and solution techniques provides a "bag of tricks" we can use for system design purposes.

To facilitate the development of this catalog, Table 6.2 lists a summarized bibliography for each managerial function. The bibliography is based entirely on books and recent survey papers, which in turn contain comprehensive lists of references for each subject area. The focus is primarily on normative, model-based approaches. The bibliography is not intended to be exhaustive. Our aim has been to suggest a limited number of useful publications in each managerial function. Many of the survey papers (Cohen [6.15], Gabbay [6.26], Golovin [6.28], Hax [6.35], and Karmarkar [6.55]) have been prepared by an MIT team working under a research effort sponsored by the Office of Naval Research.

6.2.5 The Product Structure Framework. Another fundamental input to the determination of a logistics support system is the nature of the product structure of the firm. The most general product structure can be represented by Figure 6.1. The logistics activities associated with the various elements of the product structure can be grouped into three major categories: purchasing, production (including fabrication and assembly), and distribution.

Purchasing deals with the procurement of raw materials, tools, supplies, maintenance parts and purchased parts,

Table 6.2 Bibliography by Major Logistics Function

Topic	Books	Survey Papers
Forecasting	Brown [6.9]-Exponential Smoothing, Box and Jenkins [6.8] and Nelson [6.62]-Time Series, Theil [6.70]-Econometrics, Draper and Smith [6.19]-Regression, Schlaifer [6.64]-Subjective Assessment	Chambers, Mullich and Smith [6.13]
Facilities Design	Eilon, Watson-Gandy, and Christofides [6.21], Francis and White [6.25] Scott [6.65]	Cohen [6.15], Atkins and Shriver [6.4], Geoffrion [6.27], Francis and Goldstein [6.24]-Bibliography
Aggregate Production Planning	Brown [6.10], Buffa and Taubert [6.11], Eilon [6.20], Elmaghraby [6.22], Groff and Muth [6.29], Hanssmann [6.32], Holt et al. [6.39], Magee and Boodman [6.57], Mize et al. [6.58], Naddor [6.61], and Starr [6.69].*	Hax [6.35] Silver [6.68] Elmaghraby [6.23]-Bibliography
Inventory Control and Purchasing	Arrow, Karlin and Scarf [6.3], Hadley and Whitin [6.31], Wagner [6.73]	Gross and Schrady [6.30] Golovin [6.28] Veinott [6.72]

Table 6.2 Bibliography by Major Logistics Function (continued)

Topic	Books	Survey Papers
Production Scheduling	Conway, Maxwell and Miller [6.17], Muth and Thompson [6.60]	Golovin [6.28], Day and Hottenstein [6.18]
Distribution	Bowersox [6.7], Ballou [6.5], Magee [6.56]	Clark [6.14], Karmarkar [6.55]
Vehicular Scheduling	Eilon, Watson-Gandy and Christofides [6.21]	Gabbay [6.26]
Information Processing	Blumenthal [6.6], Murdick and Ross [6.59]	Hax [6.34]

*These books cover most of the other logistics functions; in particular, they cover inventory control, production scheduling, purchasing, and information processing decisions.

subassemblies and finished products. These are the basic inputs to the production and/or distribution activities obtained from external sources. Production encompasses the fabrication and assembly operations and represents the process of converting raw materials into finished goods. It is useful to distinguish between fabrication and assembly activities since the manufacturing characteristics and the managerial decisions associated with these two processes are quite different. Table 6.3 provides a gross comparison of fabrication and assembly operations. It is worth noting that the fabrication and assembly activities can overlap in a given manufacturing situation. The resulting production process could then consist of several stages with alternating fabrication and assembly operations. Job

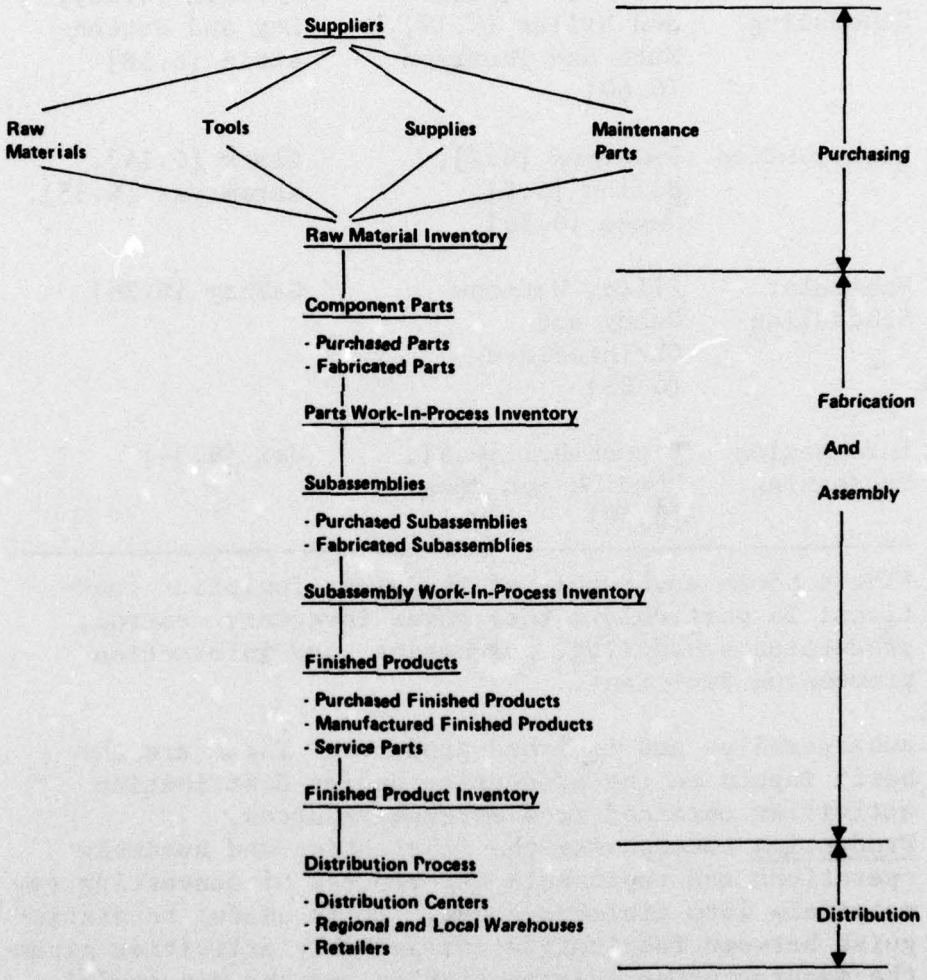


Figure 6.1 - Product structure.

Table 6.3 Production Process Taxonomy

Differentiating Factors	Fabrication	Assembly
Type of Process	Intermittent	Continuous
Process Configuration	Job Shop	Assembly Line
Locus of Decision-Making	Decentralized	Centralized
Equipment	Flexible	Specialized
Labor	High Skilled	Low Skilled
Scheduling	Dispatching	Permanent Design

shops are further classified into open and closed job shops. In an open job shop, all products are made to order, which means that normally no inventory is accumulated (for example, a repair-maintenance shop). In a closed job shop, routine demand for finished products can be forecast, production run in batch sizes, and inventories carried.

Distribution deals with the allocation of finished products from production or supply centers to customers or intermediate destinations. The product structure and the associated logistics activities have important implications that we now will examine.

6.2.6 Implications of the Product Structure Framework: A Classification of Logistics Systems and Issues of Aggregation of Information. Obviously, not every firm possesses every element of the product structure identified above nor is it engaged in all of the activities listed. For example, most retail or wholesale firms are concerned only with purchasing activities; many production plants have a production process with exlusively fabrication operations; most manufacturing firms are not involved directly with distribution operations, and so on.

In order to prescribe a logistics support system to an organization, it is imperative to resolve two basic issues related to the product structure.

What elements are present in the product structure of the firm, and what degree of complexity is involved in making the corresponding purchasing, production, and distribution decisions? To answer this question we will propose a classification of logistics systems based on the product structure and the logistics activities undertaken by the firm.

How can we aggregate the final items produced and distributed by the firm, which may amount to several thousands, so that proper planning and control procedures are defined? This question will be answered by suggesting a classification scheme based on the product structure characteristics.

Classification of Logistics Systems. To deal systematically with the first issue raised, it is useful to classify the logistics systems into four categories that are suggested by the types of activities the firm is engaged in.

Pure Inventory Systems. These systems are intended to support decisions regarding the replenishment of inventories for individual items. The decision rules associated with these systems are statistically based and specify for each item an order point (that determines when the item should be ordered) and an order quantity (that determines how much to order). Each item is treated in complete isolation except, perhaps, to allow for simple quantity discounts resulting from joint ordering of items, and to account for simple constraints reflecting storage, financial, or other limitations.

Pure inventory systems are normally applicable to raw material purchasing decisions, and retail or wholesale activities, where items are purchased from outside vendors. Pure inventory systems also can be used to control the production of finished goods in very simple manufacturing environments that are not affected by significant fluctuations in demand requirements and where ample production capacity is available. These conditions rarely are met in most production environ-

ments and therefore pure inventory systems are basically used to support only purchasing decisions.

Practical and theoretical developments of pure inventory systems have received a great deal of attention. Today, most computer manufacturers offer a wide variety of pure inventory software packages. References [6.40], [6.41], [6.44] and [6.47] describe some of the IBM packages in this field.

Production-Inventory Systems. These systems apply to situations where the firm manufactures the finished products internally rather than procuring them externally. Production activities imply allocation of scarce resources among items that are competing for capacity utilization. This means that manufactured items cannot be controlled by simple order point-order quantity rules that ignore item interactions. The rules have to be modified to take into account higher level decisions regarding capacity and labor constraints. The resulting methodology varies significantly with the type of production process involved in the manufacturing activities. In particular, a fabrication or intermittent process has to be controlled in quite a different way from an assembly or continuous process.

Obviously, the development of production-inventory systems is a much more complex task than the design of a pure inventory system. Applications based on an integrative hierarchical approach have been conducted by Hax and Meal [6.38] for a fabrication closed job shop process, by Armstrong and Hax [6.2] for a fabrication open job shop process, and by Hax [6.33] for a continuous production process. A computer-based production-inventory system implemented at MIT is documented in [6.37].

Distribution-Inventory Systems and Production-Distribution-Inventory Systems. Support systems for distribution activities are the most difficult to design. As we have indicated above, distribution involves the allocation of available inventory (either manufactured internally or purchased from outside vendors) among a set of stocking points located within a possibly complex network. In practice, sound applicable management science support is not available for these types of decisions. These are two different procedures that normally are used to deal with these decisions. In the

first one, referred to as a push system, the allocation of inventory is decided centrally taking into account all the distribution requirements and stock availabilities. Mathematical programming models are instrumental in supporting push systems. In the second procedure, referred to as a pull system, the individual warehouses independently generate requests for inventory replenishment based on their own inventory status. Statistically based inventory models have been associated with pull system decisions. Karmarkar [6.55] surveys the most important practical contributions in this area and suggests a hierarchical approach for the design of support systems involving distribution activities.

Needless to say, a given firm could have more than one of these logistics systems in operation. In many instances purchasing decisions can be handled by means of a pure inventory system, while a production-inventory system may be supporting the control of manufactured items.

Aggregation of Items. Another issue that can be resolved by analyzing the product structure of the firm is related to ways in which individual items can be aggregated into families or types for purposes of aggregate capacity planning. Proper aggregation of information is of crucial importance in the design of hierarchical logistics systems. We have found in our previous work (Hax and Meal [6.38] and Hax and Golovin [6.37]) that it is useful to consider three categories of item aggregation.

Items (also called stock-keeping units or SKUs) refer to the highest degree of specificity regarding the manufactured or purchased products. Items differ in terms of packaging characteristics, size, color, dimensions, and so on.

Families of items are formed by items sharing a common manufacturing setup cost or a common purchasing ordering cost. Economies of scale are accomplished by a joint replenishment of items belonging to the same family.

Product types constitute aggregations of items or

families of items whose production quantities are going to be determined by an aggregate plan. When models are used to support these aggregate planning decisions, the parameters used to describe the nature of the planning effort determine the common characteristics that items or families of items should possess in order to become members of a given product type. Normally, items or families of items belonging to a product type should have similar production, inventory holding, and manpower costs, similar productivity rates, and similar demand patterns.

As we indicated above, a pure inventory system deals with individual items. However, most systems involving production and distribution activities require items to be aggregated into families and types.

6.3 The Evolution of Computer Based Logistics Support Systems

Since the early developmental stages of computer technology and management sciences an important amount of work has been devoted to the applications in the field of logistics. For the ordinary manager, the most critical element that determines the feasibility of using these modern approaches is the availability of off-the-shelf computer systems that can be directly applied to his own logistics decisions, with very little or no additional programming effort on the user's part. This fact has been clearly understood by computer manufacturers and software firms which, through the last twenty years, have been engaged in a continuous effort to design and market a wide variety of computer based logistics support systems.

Although these efforts have produced an undeniably significant impact on today's business activities, it also is unquestioned that all the potentials of the new management technology have not yet been totally achieved. We will now review the evolution of the design concepts behind these computer systems in order to assess their strengths and weaknesses and to propose changes in the design criteria that might lead to a greater impact of computers and management science in the field of logistics.

Throughout these discussions, we will concentrate our

attention mostly on the work performed by IBM Corporation. We have selected IBM as the target for our analysis in order to limit the scope of the study and because IBM is the leading institution in the computer industry. By doing this, we do not intend to disregard the importance of the contributions made and that continue to be made by other computer manufacturers.

6.3.1 First Phase: Disconnected Computer Packages. Most of the work done in the design of computer based logistics systems has been oriented toward the development of isolated computer packages to support decisions in a specific managerial functional area. Typical examples are packages in inventory control, forecasting, capacity planning, and shop floor control. Table 6.4 lists some of the packages that are available to assist managers in each major logistics functional area.

The majority of these packages are sound from a methodological point of view, and they represent genuine opportunities for improving management practices in their corresponding functional areas. However, they have a number of limitations which have made their use rather restricted and their impact rather discouraging. Among the most important limitations are the following.

Lack of integration. By the nature of their design, the packages are circumscribed to an isolated area of application. The packages are disconnected among themselves, ignoring possibly greater payoffs from a broader integrative approach. They do not support effectively the overall management decision process. Coordination among the functional areas is not attained, and the applications normally lead to suboptimal solutions.

Excessive rigidity. Most of the packages have a very rigid design. Any change from its original conception is hard to implement. Modifications of the internal calculations performed by the package intended to adapt it for a slightly different application than the one originally intended, are difficult or impossible to execute. Changes in input and output specifications, and date and file management also are hard to accomplish. This clearly limits the

Table 6.4 Examples of Software Packages Available to Support Logistics Functions

Topic	Computer Package
Forecasting	Statistical Package for the Social Sciences [6.63]
Facilities Design	Mixed Integer Programming Packages:* MPSX [6.54], UMPIRE [6.71], OPHELIE [6.12] Special Packages: MULTICOM [6.66], POLIGAMI [6.67], FLAC [6.27], CAPFLO [6.27] Survey Paper: Geoffrion [6.27]
Aggregate Production Planning	IBM Capacity Planning, Finite and Infinite Loading [6.46] IBM Requirements Planning [6.45]
Inventory Control and Purchasing	IBM Retail IMPACT [6.40], IBM Wholesale IMPACT [6.41], Inventory Control Package [6.44], COGS [6.47]
Production Scheduling	IBM Bill of Materials Processor [6.49], IBM Shop Floor Control [6.51], IBM Dispatching and Job Control [6.50]
Distribution	IBM Wholesale IMPACT [6.41]
Vehicular Scheduling	IBM Vehicular Scheduling Program [6.42]
Information Processing	IBM Generalized Information System [6.48]

*Also applicable for aggregate production planning.

usefulness of the packages. In addition, they seldom can be called as subroutines by a main program, which makes the integrative issue discussed above even more pressing.

Excessive sophistication. Most packages demand a high level of sophistication on the part of the user to understand fully the basic assumptions imbedded in the package design, the ways to interact with the system, and the decisions the system can and cannot support. This has led to numerous misapplications of some of the packages, causing severe problems in some organizations. A typical example is the use of pure inventory control systems to manage production operations, ignoring interactions among items and constraints imposed by limited production capacity. The absence of diagnostic capabilities that might indicate whether or not a given package is suitable for supporting a specific decision is an inherent weakness of most logistics packages.

Finally, a significant effort is required usually to initiate, operate, and maintain the systems. This limitation is particularly critical when organizational turnover removes the personnel familiar with the system and new, uninitiated personnel have to assume control of the system operation.

6.3.2 Second Phase: The Integrative Approach. In order to correct the first of the weaknesses mentioned above, the lack of an integrative approach, IBM has developed the Production Information and Control System (PICS), [6.43], and, more recently, the comprehensive Communication Oriented Production Information and Control System (COPICS), reported in eight volumes in Reference [6.52]. This work represents a formidable effort to unify the information requirements of the firm and to support the logistics decisions within an integrated framework.

PICS and COPICS describe an overall production decision support system where the logistics managerial functions are put into a proper integrative context. Many of the disconnected packages reviewed above become now elements of a well-coordinated overall production system.

Although the basic ideas introduced in PICS and COPICS

AD-A050 798 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 15/5
MODERN TRENDS IN LOGISTICS RESEARCH. PROCEEDINGS OF A CONFERENC--ETC(U)
1976 W H MARLOW N00014-75-C-0729
UNCLASSIFIED NL

2 OF 5
AD A050798



have been partially applied in some organizations, it is fair to say that they are mostly conceptual approaches which are still waiting for an efficient way to be implemented. In fact, there is no software system available to apply the PICS or COPICS approach. Moreover, lack of diagnostic capabilities and excessive rigidity and sophistication still persist as potential limitations to implement these concepts.

6.3.3 Third Phase: The Customized Approach. Two important contributions have been made recently to design systems that are more responsive to specific user needs and to shorten the time required to implement the system. These two approaches are the IBM System/3 Application Customizer Service [6.53], and the Distribution System Simulator (DSS) [6.16].

In both of these systems the user fills out a questionnaire. The questionnaire provides a framework in which the user can think about his problem: his requirements, the data needed to choose a method of operation, and his choice of procedures. The customizer produces as output programs for functions such as order filling, billing, accounts receivable, and sales analysis. The simulator produces as output a simulation model of the user's distribution system.

One hears conflicting reports about the success of the customizer; apparently, although it is of some utility, it leaves much to be desired. The current version of the customizer, which is limited to deal with pure inventory control decisions, does not provide an effective diagnostic tool to identify what are the user's inventory problems, and to prescribe the most appropriate support system to deal with these problems. In fact, the customizer postulates that an IMPACT like inventory control system [6.40, 6.41] will serve to satisfy the user's needs and the sole objective of the questionnaire becomes the acceleration of the implementation of that system. Moreover, only one version of the customizer is available for all industries. This means that knowledge about industry-specific terminology, practices, and procedures is missing from the questions and the resulting programs.

An approach to the modelling issue has been made with the simulator. In our view, its strengths and

weaknesses are similar to those of the customizer (Hax [6.36]) and are as follows.

> The idea of facilitating the development of a customized model to serve specific user's needs, presented via a questionnaire, is a viable and interesting one, very much worth exploring. The overall concept of DSS is attractive and represents a step in the right direction to help an educated user in formulating a model-based approach to improve the quality of the decision-making in a logistics environment.
>
> The organization of DSS (in terms of an external answer sheet, and internal source library, decision table, editor and output generator) is very well accomplished and can serve as a prototype for an effort of this kind.
>
> The specific models being produced and the underlying simulation approach that dominates the design of DSS are, in our opinion, fairly ineffective. We believe DSS fails to be of assistance in most of the major questions posed to the manager of a logistics system, and we would find it very hard to use DSS intelligently in a practical situation, except in resolving some general policy issues associated with the logistics system.

Specifically, the strongest limitations of DSS are the following.

> It fails to support plant and warehouse location decisions, as well as decisions regarding expansions or improvements in the production and distribution facilities. By approaching the problem only from a simulation point of view, instead of using an optimization approach, DSS is hopeless in this respect.
>
> It ignores the production process and the interaction of the production process with a complex distribution process. DSS treats the manufacturing plants as a source of unlimited inventory, which is overly simplistic.

It does not provide an integrative approach to the logistics decision process. Essentially, DSS treats each stocking point as if it were independent of the rest of the system. The difficult problem faced in a multi-level, multi-item distribution situation is the optimum allocation of the total available inventory among the various stocking points. Basic issues to be resolved are: where to stock a given item and what strategies to use in the stock allocation, replenishment and transshipment processes. A simulation approach to deal with these issues seems highly inadequate to us.

In its present form, DSS seems inappropriate to deal with the day-to-day decisions at the operational level, which require much more detailed information.

6.4 A Proposed Approach for Logistics System Design

In this section we list the essential characteristics we feel a logistics decision support system should possess and we suggest a way to facilitate implementation.

6.4.1 Essential Characteristics of an Effective Logistics System. Having reviewed and criticized the state of development of logistics systems, we will proceed to outline the basic characteristics that we feel an ideal system to support logistics decisions should possess.

Hierarchical Structure. The distinctive nature of the decisions involved in the logistics process--investment in new resources, allocation of existing resources, and day-to-day implementation issues--makes it mandatory to adopt an integrative and hierarchical approach for the overall logistics system design. The motivation for this approach was provided in Sections 6.1 and 6.2.2 and will not be repeated here. The basic questions to be resolved when designing a hierarchical system are the following.

How to partition the decision process into modules or subproblems that properly represent the various levels of decision-making in the organizational structure

How to aggregate and disaggregate the information through the various hierarchical levels

How to solve each of the subproblems identified by the partitioning procedure

How to determine what linking mechanisms should be used among the subproblems

How to evaluate the overall performance of the system, particularly with regard to issues of suboptimization introduced by the hierarchical design

These questions are not easy to answer. Unfortunately, there is very little theoretical and empirical work that can be used as a practical guide to hierarchical design. Clearly, the questions cannot be addressed in isolation since they strongly interact with one another. Some factors that have to be taken into consideration are the following.

The organizational structure of the firm that establishes the hierarchical breakdown of responsibilities identifies the decision makers the system is intended to support and provides the basis for a preliminary decomposition of the overall decision process.

The nature of the resulting subproblems, which suggests the methodology that might be applicable to solve each of the system modules. Naturally, it is preferable to define subproblems that lend themselves to easy and effective solutions. For this stage of the problem the taxonomy presented in Sections 6.2.3 and 6.2.4 can be of assistance to the system designer.

The nature of the product structure, which is helpful in identifying ways in which information regarding individual items can be aggregated into families and product types. This subject was analyzed in Section 6.2.6.

The classification of the logistics subsystem according to the types of activities undertaken by the firm. Such a classification was suggested in Section 6.2.3.

The degree of interaction and transfer of information

from each of the hierarchical levels of the system. An effective design should facilitate the specification of the constraints that higher level decisions imposed on the lower hierarchical echelons, and the control feedback that is transferred from the lower to the higher level decisions. In addition, the feasibility of disaggregation of information should be guaranteed throughout the process, and measures of performance should be available to assess the overall quality of decision-making.

Much research needs to be allocated in order to obtain a satisfactory answer to these issues. Meanwhile, the task of hierarchical system design is an art that requires great doses of pragmatism and experience to be accomplished properly. Few practical applications of hierarchical systems have been reported in the literature. From recent work conducted at MIT, we can cite publications by Hax [6.33] dealing with a continuous manufacturing process, Hax and Meal [6.38] addressing the use of hierarchical systems in a batch processing environment, and Armstrong and Hax [6.2] describing an application for a job shop activity.

Optimization Approach. Whenever possible, an optimization approach to decision-making should be adopted. This is particularly important at the higher echelons of the hierarchical process where the broad scope of the decisions considered and the large number of interactions characteristic of these decisions make a simulation or heuristics approach particularly vulnerable. If arbitrary rules are set up, based on heuristics principles, an effort should be made to evaluate the degree of suboptimization these rules can introduce. Whenever the presence of severe uncertainties and complex precedence relationships in short time horizons make the use of simulation techniques unavoidable, it is advisable to utilize a higher level optimization model to generate the alternatives the simulation model should test. For an example of a combined optimization and simulation approach in a hierarchical system, see Armstrong and Hax [6.2].

Operational and Testing Modes. The system should be designed to operate efficiently in two different modes.

Operational mode, aimed at supporting the strategic, tactical and operational decisions in a routine and systematic fashion. This mode calls for a careful definition of the frequency in which each element of the system is to be run, the updating frequency of the parameters and files involved, and the frequency of report generation.

Testing mode, which allows for "what if" questions to be answered, and which permits a comparison of the proposed system's performance against past practices of a given organization (when using the system with historical data) for purposes of validating the preliminary design proposals.

Customized System Design. The diagnosis of the user's needs, the efforts associated with system design and implementation, and the necessary work related to system maintenance and updating can be facilitated greatly by the development of an interactive questionnaire that allows for direct input from the user. The motivation for this approach was presented in Section 6.3.3. The questionnaire, to be answered by an educated user, enables one to identify systematically the physical elements of the logistics system, the nature of the product structure, the constraints imposed on the use of the logistics system, the type of decisions the user has to make, and the economic forces that define the effectiveness of these decisions. The answers proposed to the questionnaire are processed by the system to provide preliminary cost benefit analyses and, subsequently, to suggest specific forms of system support. The approach conceived by Connors et al. [6.16] constitutes an important application of these concepts and signals the future trend for logistics system design.

Responsiveness to Management Interaction. Both in the design as well as in the operational phase, the system should provide extensive management interaction. Obvious instances where interactions should be built explicitly into the system are: at the questionnaire stage; after the development of any forecasts; and after decisions are being made in any given level of the hierarchical system (particularly at the higher levels) and are about to be transferred to the immediately lower

level. Tracking signals should be incorporated into the system, to help detect unacceptable large forecast errors and seemingly abnormal deviations from expected results.

Flexibility. The system should be easy to expand and flexible in terms of file design and retrieval, and output generation. The problem of transferability of the system to any hardware or computer operating system, although important from a practical point of view, can be ignored at an early stage of the system development.

6.4.2 The Design Approach. We believe that an orderly design of a logistics support system can be accomplished best by following a three-phase approach. These three phases might have to be repeated more than once prior to the successful completion of the design, or the reprocessing of a given phase might be advisable prior to continuing to the next one.

Phase I. Preliminary Diagnosis. The objective of this phase is to obtain an initial assessment of the impact that a model-based logistics system will have in the organization and to identify the basic characteristics of such a system. First, descriptive information is collected from the user, via the interactive questionnaire, to define the nature of the physical logistics system and its associated decision-making process. After this stage is accomplished, a representative data sample is gathered to validate some of the user's answers and to provide the basis for a preliminary cost-benefit analysis of the expected logistics system performance. An overall diagnosis is prepared, and the major logistics subsystems of the hierarchical approach are identified.

Phase II. Detailed System Design. This phase is aimed at producing a detailed design of the logistics support system. Based on the preliminary design concepts arrived at in the previous phase, a set of more specific and detailed questions is posed to the user in order to proceed to a final system recommendation. More detailed data are obtained from the user to test and validate the prescribed system. Also, final cost-benefit analyses are prepared, and a schedule for system implementation is suggested.

Phase III. System Implementation. The objective of this

phase is to facilitate and accelerate the process of system implementation. Particularly, initiation procedures are developed and data management activities are established. An important part of this effort is the specification of the input and output programs.

A similar three-phase approach for the design of production and distribution systems was proposed by Wagner [6.75].

6.4.3 Further Work. Working toward the fulfillment of the goals specified above, we have established at MIT a research project to build a system to generate logistics support systems. There are three dimensions of this work that we are studying simultaneously. These are the managerial, methodological, and computer dimensions of system development.

Managerial Dimensions. We are undertaking some field work to identify practical situations where our current work can be tested, validated and evaluated. The issues to be explored in this respect are the following.

Involvement of managers and consultants to test the quality and responsiveness of the questionnaire

Impact of the organizational structure in the development of the hierarchical system

Problems of man-machine interaction

Further refinements based on empirical testing

Specific characteristics that should be considered in the system for various types of industries

Issues of implementation

Methodological Dimensions. The central issues to be explored with regard to the methodological research are how to perform the system partitioning and how to measure the overall system performance. Specific topics to be investigated are the following.

Use of large-scale system theory in the partitioning problem

Understanding the issues of suboptimization, the use of approximations, and heuristics

Problems related to the aggregation and disaggregation of information

Formulation of interesting problems and subproblems that are of practical significance

Improved methods of solutions and special algorithms

Computer Dimensions. The initial computer work consists of setting up the questionnaire, the optimization subroutines, the operational testing and auxiliary programs, and the data management procedures. At a subsequent stage of our work, we intend to introduce artificial intelligence concepts aimed at producing a knowledge-based system. Primary goals in this direction are the following.

Questionnaire improvements and use of natural language programs

Data management innovations and automatic programming capabilities

Initial attempt to structure some of the subjective knowledge required in an appropriate system design

Knowledge-based application

We realize that the goals we have set for ourselves are exceedingly ambitious. Their ultimate achievement only can be accomplished by a long and intense effort. We believe, however, that our effort represents an overall encompassing approach which is needed badly in the field of logistics.

References

[6.1] Anthony, R. N. (1965). Planning and Control Systems: A Framework for Analysis. Harvard University, Graduate School of Business Administration.

[6.2] Armstrong, R. J., and A. C. Hax (1974). A

hierarchical approach for a naval tender job shop design. Technical Report 101, Operations Research Center, Massachusetts Institute of Technology (August).

[6.3] Arrow, K. J., S. Karlin, and H. Scarf (1958). Studies in the Mathematical Theory of Inventory and Production. Stanford University Press.

[6.4] Atkins, R. J., and R. H. Shriver (1968). New approaches to facilities location. Harvard Business Review 46 70-79.

[6.5] Ballou, R. H. (1973). Business Logistics Management. Prentice-Hall.

[6.6] Blumenthal, S. C. (1969). Management Information Systems: A Framework for Planning and Development. Prentice-Hall.

[6.7] Bowersox, D. J. (1974). Logistical Management. Macmillan.

[6.8] Box, G. E. P., and G. M. Jenkins (1970). Time Series Analysis, Forecasting and Control. Holden-Day.

[6.9] Brown, R. G. (1962). Smoothing, Forecasting, and Prediction of Discrete Time Series. Prentice-Hall.

[6.10] Brown, R. G. (1967). Decision Rules for Inventory Management. Holt, Rinehart and Winston.

[6.11] Buffa, E. S., and W. H. Taubert (1972). Production-Inventory Systems: Planning and Control. Irwin.

[6.12] Control Data Corporation (1970). OPHELIE/LP addendum: mixed integer capability of OPHELIE/LP system. Publication No. D0001507032.

[6.13] Chambers, J. C., S. K. Mullich, and D. D. Smith (1971). How to choose the right forecasting technique. Harvard Business Review 49 57-86.

[6.14] Clark, A. J. (1972). An informal survey of multi-echelon inventory theory. Naval Res. Logist.

Quart. 19 621-650.

[6.15] Cohen, J. J. (1973). A survey on the warehouse location problem. Working Paper 022-73, Operations Research Center, Massachusetts Institute of Technology (September).

[6.16] Connors, M. M., C. Coray, C. J. Cuccaro, W. K. Green, D. W. Low, and H. M. Markowitz (1972). The distribution system simulator. Management Sci. 18 425-453.

[6.17] Conway, R. W., W. L. Maxwell, and L. W. Miller (1967). Theory of Scheduling. Addison-Wesley.

[6.18] Day, J. E., and M. P. Hottenstein (1970). Review of sequencing research. Naval Res. Logist. Quart. 17 11-39.

[6.19] Draper, N. R., and H. Smith (1966). Applied Regression Analysis. Wiley.

[6.20] Eilon, S. (1962). Elements of Production Planning and Control. Macmillan.

[6.21] Eilon, S., C. D. T. Watson-Gandy, and N. Christofides (1971). Distribution Management: Mathematical Modelling and Practical Analysis. Hafner.

[6.22] Elmaghraby, S. E. (1966). The Design of Production Systems. Reinhold.

[6.23] Elmaghraby, S. E. (1973). Some recent developments in aggregate production planning and scheduling. OR Report 85, North Carolina State University (January).

[6.24] Francis, R. L., and J. M. Goldstein (1974). Location theory--a selective bibliography. Operations Res. 22 400-410.

[6.25] Francis, R. L., and J. A. White (1974). Facility Layout and Location: An Analytical Approach. Prentice-Hall.

[6.26] Gabbay, H. (1974). An overview of vehicular scheduling problems. Technical Report 103, Operations Research Center, Massachusetts Institute of Technology (September).

[6.27] Geoffrion, A. M. (1974). A guide to computer-assisted methods for distribution systems planning. Working Paper No. 216, Western Management Science Institute, University of California, Los Angeles (June).

[6.28] Golovin, J. (1973). A survey on the inventory control--detailed scheduling problem. Technical Report No. 84, Operations Research Center, Massachusetts Institute of Technology (September).

[6.29] Groff, G. K., and J. F. Muth (1972). Operations Management: Analysis for Decisions. Irwin.

[6.30] Gross, D., and D. A. Schrady (1975). A survey of inventory theory and practice. Chapter 11, this volume.

[6.31] Hadley, G., and T. M. Whitin (1963). Analysis of Inventory Systems. Prentice-Hall.

[6.32] Hanssmann, F. (1962). Operations Research in Productions and Inventory Control. Wiley.

[6.33] Hax, A. C. (1973a). Integration of strategic and tactical planning in the aluminum industry. Working Paper OR 026-73, Operations Research Center, Massachusetts Institute of Technology (September).

[6.34] Hax, A. C. (1973b). Planning a management information system for a distributing and manufacturing company. Sloan Management Review 14 85-98.

[6.35] Hax, A. C. (1974a). Aggregate capacity planning--a survey. Technical Paper 027-73, Operations Research Center, Massachusetts Institute of Technology (April).

[6.36] Hax, A. C. (1974b). A comment on the

distribution system simulator. Management Sci. 21 233-236.

[6.37] Hax, A. C., and J. Golovin (1974). A computer based production planning and inventory control system. Technical Report 102, Operations Research Center, Massachusetts Institute of Technology (September).

[6.38] Hax, A. C., and H. C. Meal (1975). Hierarchical integration of production planning and scheduling. in M. A. Geisler (ed.) Logistics. North Holland/TIMS. Studies in the Management Sciences. 1 53-69.

[6.39] Holt, C. C., F. Modigliani, J. F. Muth, and H. A. Simon (1960). Planning Production, Inventories and Work Force. Prentice-Hall.

[6.40] IBM Corporation (1965). Retail IMPACT (Inventory management program and control techniques). Application Descriptions.

[6.41] IBM Corporation (1967). Wholesale IMPACT. Advanced Principles and Implementation Reference Manual.

[6.42] IBM Corporation (1968). Vehicle scheduling program. Application Description Manual H20-0464.

[6.43] IBM Corporation (1969). The production information and control system (PICS).

[6.44] IBM Corporation (1970a). OS/360 Inventory control. Application Description Manual.

[6.45] IBM Corporation (1970b). OS/360 Requirements planning. Application Description Manual.

[6.46] IBM Corporation (1970c). Capacity planning, infinite and finite loading. Application Description.

[6.47] IBM Corporation (1970d). Consumer goods system (COGS). Application Description Manual.

[6.48] IBM Corporation (1970e). System/360 generalized information system (Basic). Application Description

Manual.

[6.49] IBM Corporation (1971). System/360 bill of materials processor. Application Description.

[6.50] IBM Corporation (1972a). Management operating system--dispatching, and job and cost reporting detail. Publication E20-0062.

[6.51] IBM Corporation (1972b). Shop floor control with IBM System/360. Publication E20-0173.

[6.52] IBM Corporation (1972c). Communications oriented production information and control system (COPICS).

[6.53] IBM Corporation (1972d). Application customizer service, application programming service, System/3, model 6. *Sales and Distribution Questionnaire.*

[6.54] IBM Corporation (1973). Mathematical programming system extended, mixed integer programming. Manual SH20-0908.

[6.55] Karmarkar, U. S. (1974). Multilocation distribution system: a survey. Technical Report 104, Operations Research Center, Massachusetts Institute of Technology (September).

[6.56] Magee, J. F. (1968). Industrial Logistics. McGraw-Hill.

[6.57] Magee, J. F., and D. M. Boodman (1967). Production Planning and Inventory Control. McGraw-Hill.

[6.58] Mize, J. H., C. R. White, and G. H. Brooks (1971). Operations Planning and Control. Prentice-Hall.

[6.59] Murdick, R. G., and J. E. Ross (1971). Information Systems for Modern Management. Prentice-Hall.

[6.60] Muth, J. F., and G. L. Thompson (1963). Industrial Scheduling. Prentice-Hall.

[6.61] Naddor, E. (1966). Inventory Systems. Wiley.

[6.62] Nelson, C. R. (1973). Applied Time Series Analysis. Holden-Day.

[6.63] Nie, N., D. H. Bent, and C. H. Hull (1970). Statistical Package for the Social Sciences. McGraw-Hill.

[6.64] Schlaifer, R. (1969). Analysis of Decisions Under Uncertainty. McGraw-Hill.

[6.65] Scott, A. J. (1971). Combinatorial Programming, Spatial Analysis and Planning. Harper and Row.

[6.66] Service in Informatics and Analysis, Limited (1970a). MULTICOM User's Manual. 23 Lower Belgrave Street, London, SW1W ONW.

[6.67] Service in Informatics and Analysis, Limited (1970b). POLIGAMI User's Manual. 23 Lower Belgrave Street, London, SW1W ONW.

[6.68] Silver, E. A. (1967). A tutorial on production smoothing and work force balancing. Operations Res. 15 985-1010.

[6.69] Starr, M. K. (1972). Production Management-Systems and Procedures. Prentice-Hall.

[6.70] Theil, H. (1966). Applied Economic Forecasting, North-Holland.

[6.71] Univac Division Sperry Rand Corporation (1970). UMPIRE (Unified mathematical programming system incorporating refinements and extensions). User's Guide, Publication S00037-00-00.

[6.72] Veinott, A. F., Jr. (1966). The status of mathematical inventory theory. Management Sci. 12 745-777.

6.73] Wagner, H. M. (1962). Statistical Management of Inventory Systems. Wiley.

[6.74] Wagner, H. M. (1972). A manager's survey of inventory and production control systems. Interfaces 2 31-39.

[6.75] Wagner, H. M. (1974). The design of production and inventory systems for multifacility and multi-warehouse companies. Operations Res. 22 278-291.

Part III

PRODUCTION, SCHEDULING, AND FACILITY LAYOUT

Chapter 7

COST AND PRODUCTION FUNCTIONS: A SURVEY*

Ronald W. Shephard
University of California, Berkeley

7.1 The Neoclassical Production Function

About the beginning of this century the notion of a production function appeared in economic analysis, to represent mathematically the technical possibilities between inputs of various goods and services (factors) and scalar output. More precisely for n factors, let $x = (x_1, x_2, \ldots, x_n)$ denote a vector of inputs per unit time of the various factors. Then a scalar valued function $\phi(x_1, x_2, \ldots, x_n)$ was taken to represent the maximal output per unit time obtainable with the vector x. This model of production is a steady state model, that is, the production possibilities represented are those where each unit of time all production activities are carried out in the same way, with no variability of products produced or goods and services used. Hence, there is no point to time dating inputs and outputs. Further the model is a net output model, that is, the scalar output rate value u of the production function $\phi(x)$ represents a final product of the production system. Intermediate products, that is, those produced in the system at one stage to be used at another for the final product, are not explicitly considered. The input vector x covers all such products, since the expenditure of resources involved is a part of the total inputs to obtain the final product.

Now, the notion as first introduced needed sharpening and the description just given is intended to assist in part in this respect. However it is necessary to clarify another point. Since $\phi(x)$ refers to "production possibilities," we must explain what is meant thereby. Some well defined technology must be had in mind. The

*The preparation of this chapter was supported by the Office of Naval Research under Contract N00014-69-A-0200-1010 with the University of California.

production possibilities may refer to technically possible arrangements which have not been realized or are not currently utilized in practice as well as those extant, not necessarily as a continuous spectrum. If convenient, one may think of the production possibilities as a finite collection of alternative and complementary processes, each with inputs used in fixed proportions to outputs (intermediate or final). The important thing is that the production function $\phi(x)$ is taken to represent the resource unconstrained possibilities. No input is bounded. The processes referred to may correspond to production arrangements using the same principles but designed for different capacities of output. However, in the model, the intensity of use is not limited. An intensity of 3 means a threefold replication of inputing the factors in this process, that is, three plants. An intensity of 3.75 means three plants operating fully each unit of time with a fourth plant used 75% of operable capacity each unit of time. The linear model for such production possibilities will be explained later on in more detail. The point I wish to make here is that the production function refers to the unconstrained technical possibilities.

Moreover, the input rates represented by the components of the input vector x are treated as independent variables. If one considers the level sets $L(u) = \{x \mid \phi(x) \geq u\}$ of the production function, that is, the input vectors x yielding at least the scalar output u, illustrated in Figure 7.1 for two factors, the input vectors x on the boundary of the level set (called ISOQUANT) show the substitutions between inputs to attain a given output rate u. In case they can be used efficiently only in a fixed combination, the set $L(u)$ takes the form of the crosshatched region spanning the dashed lines, in which case the only efficient input vector to attain the output rate u is the vector x illustrated. For any input vector x where the components are treated as independent variables, $\phi(x)$ gives the maximal value attainable with x. The set of all production possibilities represented by the production function $\phi(x)$ is not restricted to only the efficient ones for the various output rates. Efficiency is a local property. The input vector x in Figure 7.1 is

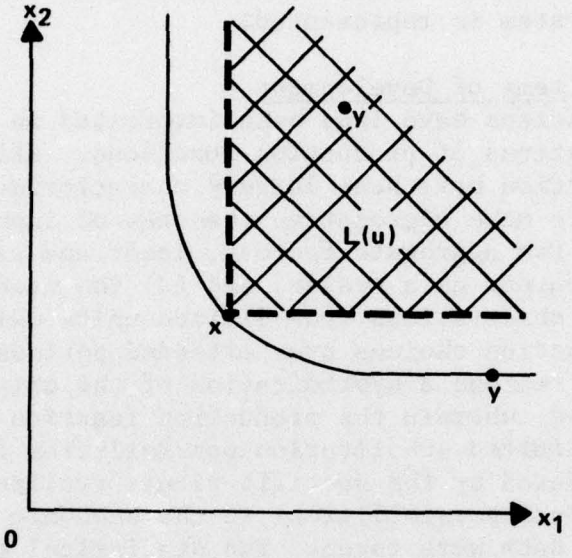


Figure 7.1 - A single efficient point x.

efficient for the output rate u because one cannot decrease any input rate and still have x belong to the level set L(u) of all input vectors which may yield the output rate u . The vectors y illustrated are not efficient for u .

When the foregoing notion of a production function is applied to various realizations of the technology as found in practice, the objects of the application involve fixed arrangements of certain limited capacities of plant and equipment. If the time perspective of the application is short and few if any alterations or additions of equipment are possible, reality cannot correspond to the model of the production function. But this is no valid objection to the theoretical construct of the production function. As Schumpeter [7.6] has said in this connection: "It is no valid objection to the law of gravitation that my watch that lies on my table does not move toward the center of the earth, though economists who are not professionally theorists argue sometimes as if it were." A proper use of the production function requires that all the limitations on inputs implied by the given facilities studied be adjoined to the production function $\phi(x)$ as constraints on the

input vectors x. Then the structure of a limited production system is represented.

7.2 Two Stems of Development

Econometricians have long been interested in statistical estimations of production functions. Efforts in this direction have been largely characterized by (a) the need to make aggregative measures of inputs and outputs, (b) two aggregate factors, labor and capital, (c) treating output as a scalar, and (d) the need to contend with observations from limited units which represent production choices over extended periods of time. For these reasons a hybridization of the original notion has evolved, wherein the production function estimated displays limited substitution possibilities for inputs as represented by the specific plants realized (among the technical possibilities) in the economic sector from which the data were taken. The statistical estimation so made may show a good descriptive relationship, but one should be chary about extrapolation for larger scales of output, unless it is clear that the mix of facilities is not to be changed. Leif Johansen [7.3] has given a clear exposition of the econometric stem of research on production functions, which I shall summarize in the next section.

The other stem is a theoretical one for the notion of a production function, to provide more adequate models which may be applied or used as perspective in applications. Starting with Leontief's [7.5] input-output model, as a linear structural relation between "industries" for estimating derived demand from a given bill of net outputs (after Quesney's Tableau Economique), and subsequent work of Koopmans [7.4] and Gale [7.2], and many individual papers along the way (see bibliography in [7.2]), the neoclassical notion of a production function has been extended to handle multiple outputs where the technical possibilities are a finite set of alternative and complementary activities. Such linear models are commonly taken to represent a total economy, after Leontief.

More recently in the theoretical stem, Shephard [7.7], [7.9], and with Färe [7.10], has developed a general set-valued production function (production correspondence) which carries the theoretical development

for steady state models to a high state of generality. Let $u = (u_1,u_2,\ldots,u_m)$ denote a vector of output rates of various goods and services that may issue from the technology and $x = (x_1,\ldots,x_n)$ denote a vector of input rates of the factors relevant to the technical possibilities. For given input vector x, $P(x)$ denotes the set of output vectors realizable from x. By a few axioms on the correspondence $x \to P(x)$ (or its inverse $u \to L(u) = \{x \mid u \in P(x)\}$) it is possible to construct a rather general theory for steady state production functions which contain the linear models as special cases. These axioms do not require assumptions that outputs and inputs are freely disposable nor that the map sets $P(x)$ and $L(u)$ be convex. In fact such assumptions are not even required for a law of diminishing returns as shown in [7.10]. The dropping of freely disposable outputs is important, because the vector u spans all net outputs whether wanted or not, as in the case of waste products.

7.3 The Econometric Production Function

In a recent book of Johansen [7.3] one finds a good account of the production function as a model for econometric studies, and a great variety of econometric studies are listed in the bibliography. Johansen distinguishes four forms of the production function.

(a) The ex ante production function at the micro level

(b) The ex post production function at the micro level

(c) The short run production function at the macro level

(d) The long run production function at the macro level

The first of these four is essentially the neoclassical production function, and the second is merely a statement for a given realized limited unit of production that inputs are proportional to output ranging from zero to some fixed bounded capacity.

The production function of main interest is the short-run macro production function. For some economic sector (under study presumably), it is assumed that there are a given number of ex post micro units available from which data may be obtained. Symbolically let

x^i = scalar output rate (same kind for all units)

$\bar{x}^i$ = output rate capacity

ξ_1^i = input of factor 1 required per unit output of ith unit

ξ_2^i = input of factor 2 required per unit output of the ith unit

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

For a production function of two factors, let v_1 and v_2 denote amounts per unit time of the two factors available. Then the short run macro production function is defined by

$$F(v_1,v_2) = \text{maximum} \sum_{i=1}^{N} x^i$$

$$\text{s.t.} \sum_{i=1}^{N} \xi_1^i \cdot x^i \leq v_1$$

$$\sum_{i=1}^{N} \xi_2^i \cdot x^i \leq v_2$$

$$0 \leq x^i \leq \bar{x}^i \quad (i = 1,2,\ldots,N)$$

which gives the maximal output over all ex post micro units of the sector, subject to the capacity constraints and the amounts per unit time v_1 and v_2 of the two factors available. This then is the physical basis for the short-run macro production function $F(v_1,v_2)$.

If one wishes to interpolate for the given ex post micro units of the sector, output obtainable from these units for various availabilities of the two factors, the function $F(v_1,v_2)$ provides the answer unless the individual agents controlling the ex post micro units do not cooperate to maximize Σx^i. The properties of this kind of production function are certainly different than those of the neoclassical, or theoretical, or ex ante, production function. For one thing, $F(v_1,v_2)$ is bounded for both v_1 and v_2 being unbounded. Also the substitution possibilities between v_1 and v_2 are more limited.

The statistical estimation of the short-run macro production function has to proceed without detailed knowledge of the input ratios ξ_1^i, ξ_2^i and the capacities $\bar{x}^i$. Often output is also not a scalar. Then one has to measure output, and the two inputs (labor and capital) by price deflated dollar values or other means of aggregate measure. Further, since it is often wished to use the production function for future projections, one has to face the problem of the vintage of the ex post micro units for the sector. By taking the macro production function as (where K and L are used for v_1 and v_2)

$$F(K,L,t) = f_1(t)\cdot G(K,L)$$

or

$$F(K,L,t) = H(K,f_2(t)\cdot L)$$

or

$$F(K,L,t) = J(f_3(t)K,L)$$

for output augmenting, labor augmenting, and capital augmenting, respectively, one may try to make statistical fits to time series data that can be predictive of what is called "technological progress." In case $F(\lambda K,\lambda L,t) = \lambda F(K,L,t)$, technological progress is said

to be Hicks neutral. Here t denotes the time at which K and L have been (are to be) applied. Mathematical functions of simple parametric form like Cobb-Douglas or the CES (constant elasticity of substitution) functions are usually taken for G, H and J. The technological progress so represented may be embodied as in the case of new equipment or disembodied (better nonembodied) as in the case of increased efficiency of management through learning or other means.

Johansen's formulation of the long-run production function involves such restrictive equilibrium assumptions that it will not be taken up here.

There are other obvious difficulties in the econometric approach to the formulation and statistical estimation of a production relationship. The use of aggregate inputs is particularly bothersome when they involve a scalar measure in dollar value terms. For a real measure a proper price deflation of dollar values in a time series is required. Even if this can be done, the estimated production relationship is conditioned by the physical output mix of the data used, and, if a substantial change of output mix is of interest for projection, the statistically estimated price deflated dollar value of output may be far from what is sought. Also, in the management of technologies, the price deflated scalar dollar value of many outputs may not be of primary interest.

All of these problems lead some of us more and more to a micro economic approach like that of the neoclassical production function, by an activity analysis or more general treatment for a production correspondence.

7.4 Generalized Neoclassical Production Functions

As indicated above the developments here are in terms of production correspondences, first linear and then general.

The linear models of Leontief, Koopmans, Gale, and others are multi-sector and refer to the total economy. The development of linear models to be given here is a formulation for the production function of a technology. For this technology we assume that there are a finite number K of processes or activities as alternatives and complements. Related to these activities there are overall n goods and services which may enter as

exogenous inputs. When an input is used by an activity, the amount required is proportional to the intensity with which the activity is applied. Let $x = (x_1, x_2, \ldots, x_n)$ denote a vector of input rates for the n exogenous inputs. Let $z = (z_1, z_2, \ldots, z_k)$ denote a vector of intensities for the k processes. Let

$$A = \begin{bmatrix} a_{11}, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22}, \ldots, a_{2n} \\ \vdots \\ a_{k1}, a_{k2}, \ldots, a_{kn} \end{bmatrix}$$

be a matrix of input technical coefficients, one row to an activity, with a_{ij} being the input of the jth exogenous input per unit intensity of the ith activity. For any given vector x the alternatives for activity intensities are given by those vectors z satisfying $zA \leq x$. Suppose there are m net products for the technology represented by the k activities. Let $u = (u_1, u_2, \ldots, u_m)$ be a vector of output rates for these m products.

At this point we must make an assumption about disposability of inputs and outputs. The simple assumption and one used by most economic theorists is: that outputs and inputs are disposable, that is, if u^0 can be produced any vector v such that $0 \leq v \leq u^0$ can also be produced, and if x^0 yields an output vector u, then any input vector y such that $y \geq x^0$ will also yield u.

With these assumptions, the output vectors u producible by an intensity vector z for activities are given by $u \leq zB$, where

$$B = \begin{bmatrix} b_{11}, b_{12}, \ldots, b_{1m} \\ b_{21}, b_{22}, \ldots, b_{2m} \\ \cdot \\ \cdot \\ \cdot \\ b_{k1}, b_{k2}, \ldots, b_{km} \end{bmatrix}$$

is a matrix of output technical coefficients with b_{ij} being the output of the jth net product per unit intensity of the ith activity.

Thus for a given input vector x, the set of all output vectors producible under free disposability of inputs and outputs is

$$P(x) = \{u \mid z \cdot A \leq x \ , \ u \leq z \cdot B\}$$

where it is understood that u, x and z are non-negative vectors. The relationship $x \to P(x)$ is a point to set production function or correspondence. In case inputs and outputs are not taken freely disposable,

$$P(x) = \{u \mid z \geq 0 \ , \ z \cdot A = \lambda x \ , \ \lambda \in [0,1] \ , \ u = \theta \cdot zB \ , \ \theta \in [0,1]\}$$

For $\lambda = 1$ and intensity z such that $z \cdot A = x$, $u = z \cdot B$ is obtainable and also any scalar fraction θzB for θ ranging from 0 to unity. Then, in case some outputs are unwanted, one can reduce them from those of $u = zB$ only by scalar reduction of the whole vector u.

For the same linear technology there is another point to set production function inverse to $x \to P(x)$, defined by $u \to L(u)$ where $L(u) = \{x \mid u \in P(x)\}$ is the set of input vectors which will yield at least the output vector u. In the two cases of free disposal and nonfree disposal of inputs and outputs

$$L(u) = \{x \mid z \cdot B \geq u \ , \ x \geq z \cdot A\}$$

$$L(u) = \{x \mid z \geq 0 \ , \ z \cdot B = (1/\theta) \cdot u \ , \ \theta \in (0,1] \ ,$$

$$x = (1/\lambda) \cdot z \cdot A \ , \ \lambda \in (0,1]\}$$

When disposable, those activity intensities yielding at least an output vector u are given by $z \cdot B \geq u$ and any input vector x such that $x \geq zA$ qualifies to yield at least u. In the nondisposable case, $zB = u$ for $\theta = 1$ so that $x = zA$ for $\lambda = 1$ will yield u, and scalar increases $(1/\lambda)zA$ of the input vector x will also yield u, that is, scalar increases of a feasible input vector x for u can be disposed to obtain u.

The foregoing linear model (production function) can be specialized to yield Johansen's type of model for the econometric short-run macro production function as follows. We imagine that the activities represent alternative means of production, blueprinted to different specific output capacity for some critical or important component of the related output vector, with the component of z taken unity at this capacity. An intensity of 3.5, say, for such an activity means a fourfold replication of the related blueprint is realized, with one of the realized facilities operating only 50% of the time at capacity during each unit of time. For the "sector" studied, let the first α activities be realized while the others are not. Then

$$0 \leq z_i \leq 1 \quad \text{for} \quad i \in \{1,2,\ldots,\alpha\}$$

$$z_i = 0 \quad \text{for} \quad i \in \{\alpha+1,\alpha+2,\ldots,k\}$$

and for given input vector x,

$$\tilde{P}(x) = \left\{ u \;\middle|\; \begin{array}{l} 0 \leq z_i \leq 1 \ , \ i \in \{1,2,\ldots,\alpha\} \\ z_i = 0 \ , \ i \in \{\alpha+1,\ldots,k\} \end{array} \quad zA \leq x \ , \ u \leq zB \right\}$$

for the realizable outputs in the sector. The correspondence $x \to \tilde{P}(x)$ is the generalization of Johansen's "short-run macro production function," in the case of m outputs. Disposability of inputs and outputs was

used to be analogous. It is not possible to speak of a maximal output, nor any need to do so. Moreover the model $x \to \tilde{P}(x)$ is in micro form rather than macro, since the use of only two inputs (usually "capital" and "labor") is fictitious and no means is given to aggregate micro inputs into two macro inputs in the econometric formulation.

The meaning of short-run is not clear and the distinctions drawn by Johansen vis-à-vis the ex post production functions are too coarse. In a span of four to five years new and larger scale production facilities can be constructed to supplement the existing facilities, as was done in the U.S. economy during World War II. If this span of time is short, one can construct a "short-run plus" production function by applying only the constraints $0 \leq z_i \leq 1$, $i \in \{1,2,\ldots,\alpha\}$ and defining $\tilde{P}(x)^+$ by

$$\tilde{P}(x)^+ = \left\{ u \;\middle|\; \begin{array}{l} 0 \leq z_i \leq 1 \text{ , } i \in \{1,2,\ldots,\alpha\} \\ z_i \geq 0 \text{ , } i \in \{\alpha+1,\ldots,k\} \end{array} \quad zA \leq x \text{ , } u \leq zB \right\}$$

in the disposable case.

Thus one may see the generality, as it is, of the linear model of production. At this point it is useful to dwell more specifically on the assumptions and limitations of this model. Not all of the input coefficients a_{ij} of A need be positive, since the list of exogenous inputs may contain alternative factors. But in each row of A at least one coefficient must be positive, otherwise net products can be produced without exogenous inputs. In the linear net output model $x \to P(x)$, intermediate products are not explicitly considered and each activity, as conceived, must integrate over intermediate products to relate the exogenous inputs to net outputs. In these circumstances some exogenous inputs are required for every activity. Returning to the columns of A , at least one positive coefficient must exist for each column, otherwise an exogenous input is not used by any activity and can be deleted.

For the matrix B , not all coefficients b_{ij} need

be positive. An activity (row of B) may produce only part of the net outputs represented by the output vector u, but there must be at least one positive coefficient; otherwise an activity produces no net product and should have been integrated with another activity. In each column of B there must be at least one positive coefficient; otherwise one of the net products cannot be produced and should be deleted.

The correspondence (production function) $x \to P(x)$ for the linear model has one very strongly limiting property for its use, namely: $P(\lambda x) = \lambda P(x)$ for $\lambda \in [0,+\infty)$, that is, it is homogeneous of degree +1, implying constant returns to scale of input. Reverting again to the conception of the activity structure used in discussing the "short-run macro production function," the replication of plants of different capacities, represented by intensities $z_i \geq 1$, is not seriously limiting for optimal planning of production. However, it is unrealistic to allow $0 < z_i << 1$ where $<<$ means "considerably less than." In considering the alternatives (of varying capacities for output), one should lower bound each positive intensity by unity, that is, $z_i \geq 1$, with $z_i = 0$ as the only alternative less than unity, that is, apply the constraints $z_i(z_i-1) \geq 0$ $(i=1,2,\ldots,k)$. Then the model ceases to be linear, and increasing returns to scale are possible, that is, $P(\lambda x)$ contains $\lambda \cdot P(x)$ as a proper subset.

Constructing a net output linear model ab initio for the production function $x \to P(x)$, where $P(x)$ is the set of all possible output vectors realizable with x, avoids certain details on intermediate products that ought to be considered. For this purpose, let the matrices A and B be defined as before, with an activity now possibly being one which yields only intermediate products, requiring that we extend the list of outputs to include intermediate products. Let

$$\bar{A} = \begin{bmatrix} \bar{a}_{11},\ldots,\bar{a}_{1m} \\ \bar{a}_{21},\ldots,\bar{a}_{2m} \\ \vdots \\ \bar{a}_{k1},\ldots,\bar{a}_{km} \end{bmatrix}$$

denote a matrix of intermediate product input coefficients, where $\bar{a}_{ij} \geq 0$ for all i and j, without requiring that there be a positive coefficient in each row and column. An activity may not use any of the outputs as intermediate product and some outputs may be entirely final product. An output may be both final product and intermediate product, but we do not regard a good or service as both exogenous input and intermediate product.

There are two alternative constraints which may be used for viability of the linear production model as follows.

I $z \cdot \bar{B} \geq 0$, $\bar{B} = (B-\bar{A})$

II $b_{ij} \geq \bar{a}_{ij}$ for all i , j

The second of these two is obviously the stronger, and I is preferred. Then the linear correspondence (production function) $x \to P(x)$ is given by

$$P(x) = \{u \mid zA \leq X \text{ , } u \leq z\bar{B} \text{ , } z\bar{B} \geq 0\}$$

$$= \{u \mid zA = \lambda x \text{ , } \lambda \in [0,1] \text{ , } u = \theta z\bar{B} \text{ , } \theta \in [0,1] \text{ ,}$$

$$z\bar{B} \geq 0\}$$

Hence, it is seen that the previously formulated linear net output model is not isomorphic to the net output model where intermediate products are considered

explicitly, unless the stronger assumption II is made, on account of the need to include the constraints $z\bar{B} \geq 0$ in the more general case.

For the inverse correspondence, the constraint $z\bar{B} \geq 0$ does not have to be applied, since

$$L(u) = \{x \mid z\bar{B} \geq u \ , \ x \geq zA\}$$

or

$$L(u) = \{x \mid z\bar{B} = (1/\theta)u \ , \ \theta \in (0,1] \ , \ x = (1/\lambda)zA \ , \ \lambda \in (0,1]\}$$

because u and $(1/\theta)u$ are nonnegative.

The open Leontief model is obtained from the foregoing one with intermediate products by (a) letting v denote a vector of total outputs with u being a vector for net products, (b) taking the matrix B to be square with unity along the diagonal and all other coefficients zero, and (c) taking the vector v as the intensity vector z, and expressing the matrix $\bar{A}$ as $C = [c_{ij}]$. Then the set of net output vectors obtainable with an input vector x is given by

$$P(x) = \{u \mid vA \leq x \ , \ v[I-C] \geq 0 \ , \ u \leq u[I-C]\}$$

or

$$P(x) = \{u \mid vA = \lambda x \ , \ \lambda \in [0,1] \ , \ v[I-C] \geq 0 \ , \ u = \theta v[I-C] \ , \ \theta \in [0,1]\}$$

Here $[I-C]$ plays the role of the matrix $\bar{B}$. The inverse correspondences $u \to L(u)$ are defined by

$$L(u) = \{x \mid v[I-C] \geq u \ , \ x \geq vA\}$$

$$= \{x \mid x \geq u[I-C]^{-1}A\}$$

or

$$L(u) = \{x \mid v[I-C] = (1/\theta)u \ , \ \theta \in (0,1] \ ,$$

$$x = (1/\lambda)vA \ , \ \lambda \in (0,1]\}$$

$$= \{x \mid x = (1/\sigma)\ u[I-C]^{-1}A \ , \ \sigma \in (0,1]\}$$

from which is seen the very special structure implied by the Leontief model. When inputs are disposable the input set $L(u)$ has a single efficient point $X^0 = u[I-C]^{-1}A$, as illustrated in Figure 7.2 where the set $L(u)$ is the shaded region, and in case inputs are not disposable, the set $L(u)$ is merely the ray L terminating below at x^0 .

So far we have been concerned with linear models for generalization of the neoclassical production function. All of them have the property that the input and output sets $L(u)$ and $P(x)$ are convex along with homogeneity. It is useful to look at the production functions $x \to P(x)$ and $u \to L(u)$ from a more general axiomatic structure.

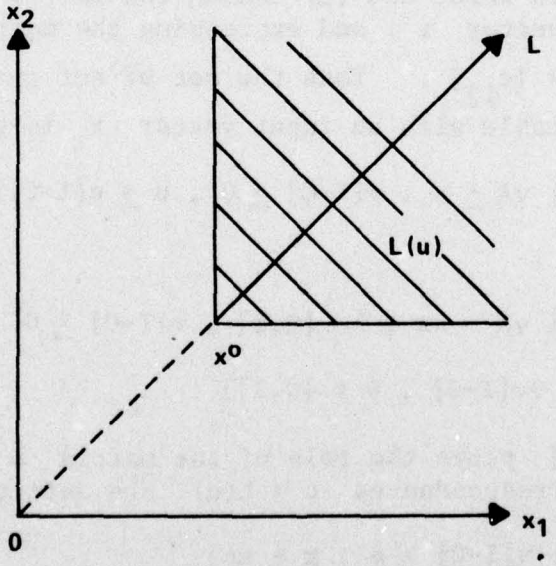


Figure 7.2 - A single efficient point x^0.

Let $x \in R_+^n$ and $u \in R_+^m$ be nonnegative vectors of exogenous input rates and net output rates respectively, and take

$P(x)$ = set of output vectors obtainable with x

$L(u)$ = set of input vectors yielding at least u

Then as axioms we may assume (see [7.9] or [7.10]) the following where semipositive vectors are nonnegative and nonzero.

P.1 $P(0) = \{0\}$

P.2 $P(x)$ is bounded for bounded x

P.3 $P(\lambda x) \supset P(x)$ for $\lambda \in [1,+\infty)$

P.4 If x is semipositive and there exists semipositive $\bar{u}$, $\bar{\lambda} \in (0,+\infty)$, such that $\bar{u} \in P(\bar{\lambda} \cdot x)$, then for all $\theta \in (0,+\infty)$ there exists $\lambda_\theta \in (0,+\infty)$ such that $(\theta \bar{u}) \in P(\lambda_\theta \cdot x)$. There exists a semipositive x such that $u \in P(x)$ for some $u > 0$.

P.5 The graph of $x \to P(x)$ is closed

P.6 If $u \in P(x)$, $\{\theta u \mid \theta \in [0,1]\} \subset P(x)$

The properties P.1 and P.2 are clearly applicable: the only possible output from a null input vector is the null output vector, and bounded inputs can only yield bounded outputs. Property P.3 is equivalent to a weak disposability of inputs, that is, if $x \in L(u)$ then $(\lambda x) \in L(u)$ for $\lambda \in [1,+\infty)$. Property P.4 is an axiom concerning attainability of output vectors, and in this connection one may note that not all exogenous inputs are necessarily needed for a semipositive or positive output vector, that is, they may be alternatives.

Property P.5 is a purely mathematical assumption to guarantee that the isoquants exist. Property P.6 is a weak assumption concerning disposability of outputs.

The properties of the inverse correspondence $u \to L(u)$ implied by those taken for $x \to P(x)$ are the following.

L.1 $L(0) = R^n_+$ and $0 \notin L(u)$ for semipositive u

L.2 For $\{|u^v|\} \to \infty$, $\bigcap_{v=1}^{\infty} L(u^v)$ is empty

L.3 If $x \in L(u)$, $(\lambda x) \in L(u)$ for $\lambda \in [1,+\infty)$

L.4 If x is semipositive and $(\bar{\lambda}x) \in L(\bar{u})$ for semipositive $\bar{u}$ and some $\bar{\lambda} \in (0,+\infty)$, the ray $\{\lambda x \mid \lambda \in [0,+\infty)\}$ intersects all input sets $L(\theta\bar{u})$ for $\theta \in [0,+\infty)$

L.5 The graph of $u \to L(u)$ is closed

L.6 $L(\theta u) \subset L(u)$ for $\theta \in [1,+\infty)$

With these few axioms (P.1,...,P.6) <=> (L.1,..., L.6) a general theory for production functions can be developed by adding one more postulate which is asymmetric. The output isoquant (production frontier) for an input vector x is defined by

$$\text{ISOQ } P(x) = \{u \mid u \in P(x),\ (\theta u) \notin P(x) \text{ for } \theta \in (1,+\infty)\},\ P(x) \neq \{0\}$$

while the input isoquant for an output vector is given by

$$\text{ISOQ } L(u) = \{x \mid x \in L(u),\ (\lambda x) \notin L(u) \text{ for } \lambda \in [0,1]\}$$

when u is semipositive and $L(u)$ is not empty, and

ISOQ $L(0) = \{0\}$.

Note that the axioms (P.1,...,P.6) <=> (L.1,..., L.6) do not imply that all output vectors $u \in R_+^m$ are feasible, that is, $L(u)$ is not empty. The efficient subset of $L(u)$ is not empty is defined by

$$E_L(u) = \{x \mid x \in \text{ISOQ } L(u) , y \notin L(u) \text{ for } x - y \text{ semipositive}\}$$

The asymmetric postulate (axiom) is

A: $E_L(u)$ is bounded for all $u \in R_+^n$

which is to state that an output vector u cannot be attained efficiently by unbounded application of any exogenous input.

With these few assumptions a general and comprehensive theory of production functions may be developed. The neoclassical production function $\phi(x)$ is obtained when u is a scalar by noting that $P(x)$ is an interval $[0,\phi(x)]$ where

$$\phi(x) = \text{Max } \{u \mid u \in P(x)\}$$

and the properties implied for $\phi(x)$ are the following.

A.1 $\phi(0) = 0$

A.2 $\phi(x)$ is bounded for x bounded

A.3 $\phi(\lambda x) \geq \phi(x)$ for $\lambda \in [1,+\infty)$

A.4 For any semipositive x such that $\phi(\lambda x) > 0$ for some $\bar{\lambda} \in (0,+\infty)$, $\phi(\lambda x) \to +\infty$ as $\lambda \to +\infty$.
$\phi(x) > 0$ for some semipositive x

A.5 $\phi(x)$ is upper semicontinuous on R_+^n

It is shown in [7.10] that P.1, P.4, P.5, P.6 and A are minimal for a weak law of diminishing returns, and

that P.1, P.2, P.4, P.5, P.6 and A are minimal for a strong law of diminishing returns. Convexity of the input and output sets $L(u)$ and $P(x)$ is not required. The significance of the law of diminishing returns in this connection is that it specifies an essential feature of production structure implied in weak form by the axioms taken for the structure of production.

The mappings $x \to P(x)$ and $u \to L(u)$ may not be a convenient form for mathematical representation of the production structure. However, a convenient functional representation is available. Define

$$\Psi(u,x) = [\text{Min } \{\lambda \mid (\lambda x) \in L(u) , \lambda \in [0,+\infty)\}]^{-1}$$

$$\Omega(x,u) = [\text{Max } \{\theta \mid (\theta u) \in P(x) , \theta \in [0,+\infty)\}]^{-1}$$

referred to in [7.7] as distance functions. It is shown that

$x \in L(u)$ if and only if $\Psi(u,x) = 1$

$u \in P(x)$ if and only if $\Omega(x,u) = 1$

and the map sets of the correspondence $u \to L(u)$ and $x \to P(x)$ are defined in terms of these two functions by

$$L(u) = \{x \mid \Psi(u,x) \geq 1\}$$

$$P(x) = \{u \mid \Omega(x,u) \leq 1\}$$

The two functions $\Psi(u,x)$ and $\Omega(x,u)$ have nice regular properties and they may be used to play the role of the neoclassical production function. For one thing, $\Psi(u,\lambda x) = \lambda\Psi(u,x)$ and $\Omega(x,\theta u) = \theta\Omega(x,u)$ for $\lambda,\theta \in (0,+\infty)$. They are nonnegative, $\Psi(u,x)$ is upper semicontinuous in x and $\Omega(x,u)$ is lower semicontinuous in u. When the map sets $L(u)$, $P(x)$ are taken convex and outputs and inputs are disposable, $\Psi(u,x)$ is a continuous concave function of x and $\Omega(x,u)$ is a contintuous convex function of u. Otherwise, $\Psi(u,x)$ is nondecreasing in u and $\Omega(x,u)$ is nonincreasing in x.

Further, if $x \to P(x)$ is continuous with strong disposal of outputs and $u \to L(u)$ is continuous, the function defined as $F(x,u) = (\Psi(u,x) - \Omega(x,u))$ is a joint production function, that is, for given x, $F(x,u) = 0$ defines ISOQ $P(x)$, and for given u, $F(x,u) = 0$ defines ISOQ $L(u)$. Note that unless $\phi(x)$ is continuous and strictly increasing $\phi(x) = u$ does not define ISOQ $L(u)$ where $L(u) = \{x \mid \phi(x) \geq u\}$, that is, the simple equation $\phi(x) = u$ does not necessarily define ISOQ $L(u)$ when $\phi(x)$ is the neoclassical production function.

There are certain special forms of the general production functions $x \to P(x)$ and $u \to L(u)$ which may be useful as approximations.

The first of these is that connected with homothetic structure. The input structure $u \to L(u)$ is homothetic if

$$L(u) = F^{-1}(f(u)) \cdot L_{\phi}(1)$$

where the fixed input set $L_{\phi}(1)$ is defined by

$$L_{\phi}(1) = \{x \mid \phi(x) \geq 1\}, \quad \phi(\lambda x) = \lambda \phi(x)$$

For the correct properties of the input sets so represented, it is assumed that $f(u)$ and $F(\cdot)$ have certain properties (see [7.7]). Homotheticity of input structure means that all input sets may be generated by radial expansion of a fixed set, and the set $L(u)$ may be expressed

$$L(u) = \{x \mid \phi(x) \geq F^{-1}(f(u))\}$$

and ISOQ $L(u)$ is defined by the equation

$$\phi(x) = F^{-1}(f(u))$$

Output structure $x \to P(x)$ is homothetic if

$$P(x) = F(\phi(x)) \cdot P_{f}(1)$$

where the fixed output set $P_f(1)$ is defined by

$$P_f(1) = \{u \mid f(u) \leq 1\}\ ,\ f(\theta u) = \theta \cdot f(u)$$

with the output set $P(x)$ expressed by

$$P(x) = \{u \mid f(u) \leq F(\phi(x))\}$$

and ISOQ $P(x)$ defined by the equation

$$f(u) = F(\phi(x))$$

Input and output structure can be each homothetic without the other homothetic. For homothetic input structure the expression for $L(u)$ implies

$$P(x) = \{u \mid f(u) \leq F(\phi(x))\}$$

but not $f(u)$ homogeneous. If both input and output structure are homothetic

$$F(x,u) = f(u) - F(\phi(x))$$

is a joint production function.

In what sense can homotheticity of $u \to L(u)$ be regarded as an approximation? If one is interested in a subregion of the input space R^n_+ which subtends a "small" solid angle with the origin, then homotheticity may be a good approximation. In fact, if one is interested in a single input mix x^0, $F^{-1}(f(u)) \cdot \{\lambda x^0 \mid \lambda \geq 1/\phi(x^0)\}$ will be an accurate representation of the alternatives, since $F^{-1}(f(u))$ allows enough flexibility to describe any case. Similarly homotheticity of output structure may be used as an approximation when the region of interest in the output space subtends a "small" solid angle with the origin.

Homotheticity has further significance when one considers cost and revenue. See below.

Another useful special form is that of semi-homogeneity. See [7.9]. Homogeneous production functions (correspondences) scale so that $P(\lambda x) =$

$f(\lambda)\cdot P(x)$, that is, if an input vector x is scaled by λ , the output set $P(\lambda x)$ is obtained from $P(x)$ by a scaling factor that depends only upon the scale parameter λ . It is shown that this kind of scaling can occur if and only if $f(\lambda) = \lambda^k$, where k is a positive constant. For different input vectors x of different mix, that is, $x/|x|$ is different, a scaling of the input vectors is done by the same scale factor λ^k on outputs, but on different output sets $P(x)$ in general. Homothetic output structures scale on the same output set $P_f(1)$ for scaling of an input vector x , but the scale factor $F(\phi(\lambda x))$ can be a quite generally behaving nondecreasing function.

If one assumes that $P(\lambda x) = S(\lambda, x/|x|)\cdot P(x)$, that is, the scaling is done by a function of the scale parameter λ and the input mix $x/|x|$, it is shown that this can happen if and only if

$$S\left(\lambda, \frac{x}{|x|}\right) = \lambda^{H\left(\frac{x}{|\lambda|}\right)}$$

Then the output structure $x \to P(x)$ is said to be <u>semihomogeneous</u> if

$$P(\lambda x) = \lambda^{H\left(\frac{x}{|x|}\right)}\cdot P(x)$$

Similarly, the input structure $u \to L(u)$ is semihomogeneous if

$$L(\theta u) = \theta^{G\left(\frac{u}{|u|}\right)}\cdot L(u) \ ,$$

Either structure may be semihomogeneous without semihomogeneity for the other. However, if both are semihomogeneous, the exponent function $H(x/|x|)$ is a positive constant for all $x \in L(u)$ and equal to

$$G\left(\frac{u}{|u|}\right)^{-1}$$

Semihomogeneity of structure is nontrivial under the general axioms (P.1,...,P.6) <=> (L.1,...,L.6) . However, if free disposability of inputs is applied, $H(x/|x|)$ is a positive constant for all x such that that $P(x) \neq \{0\}$, and, if free disposability of outputs is invoked, $G(u/|u|)$ is a positive constant for all u such that $L(u)$ is not empty. If convexity of input sets $L(u)$ is applied, the exponent function $G(u/|u|)$ has to be a positive constant for all connected input sets $L(u)$, and convexity of the output sets $P(x)$ requires the exponent function $H(x/|x|)$ to be a positive constant for all sets $(P(x) - \{0\})$ which are connected. The general properties (P.1,..., P.6) <=> (L.1,...,L.6) do not require the input sets and sets $(P(x) - \{0\})$ to be connected. The nonempty sets of the former may be distinct rays truncated from the origin and the sets $P(x) \neq \{0\}$ may be distinct truncated rays issuing from the origin, reflecting that the production possibilities are composed of only distinct output and input mixes.

The significance of semihomogeneity for production structure is that it permits the degree of homogeneity to depend upon output and input mix.

7.5 Cost and Revenue Functions

Letting $p \in R_+^n$ and $r \in R^m$ denote price vectors for input vectors x and output vectors u , respectively, the cost and revenue functions are defined by

$$Q(u,p) = \underset{x}{\text{Min}} \{p \cdot x \mid x \in L(u)\}$$

$$R(x,r) = \underset{u}{\text{Max}} \{r \cdot u \mid u \in P(x)\}$$

respectively. $Q(u,p)$ gives the minimal cost of obtaining the output vector u under prices p for inputs. $R(x,r)$ gives the maximal return that may be obtained from an input vector x under prices r for outputs.

Here r is not restricted to be a nonnegative vector, because one may wish to assign a negative price to certain unwanted outputs. For the linear model of production, values of these two functions are calculated by straightforward linear programs.

If input structure $u \to L(u)$ is homothetic, the cost function takes the form

$$Q(u,p) = F^{-1}(f(u)) \cdot P(p)$$

and this factorization holds if and only if the input structure is homothetic. By taking $f(u)$ as a scalar measure of outputs

$$f(u) = F\left(\frac{Q(u,p)}{P(p)}\right)$$

and the function $F(\cdot)$ which reflects returns to scale of input may be studied as a relationship between scalar measure of output rate $f(u)$ and factor price deflated cost rate $(Q(u,p)/P(p))$. The price function $P(p)$ is an index function of price level, since

$$Q(u,p) = \phi(x^*) \cdot P(p)$$

where x^* is an input vector yielding $Q(u,p)$, and $x^* \in \text{ISOQ } L(u)$ so that $\phi(x^*) = F^{-1}(f(u))$. Then, if $\phi(x)$ is taken as an index function of inputs ($\phi(x)$ is homogeneous of degree +1), the product of the two index functions $\phi(x^*)$ and $P(p)$ equals value (cost). The function $F(\cdot)$ reflects returns to scale because the isoquants of $u \to L(u)$ are defined by

$$F(\phi(x)) = f(u)$$

Similarly, if and only if the output structure $x \to P(x)$ is homothetic, does

$$R(x,r) = F(\phi(x)) \cdot \pi(r)$$

and if $\phi(x)$ is taken as a scalar measure of inputs, returns to scale as reflected by $F(\cdot)$ may be studied

by relating price deflated revenue $(R(x,r)/\pi(r))$ to the scalar measure of input $\phi(x)$. The price function $\pi(r)$ is a proper index function of output prices, since the isoquants of $x \to P(x)$ are given by $f(u) = F(\phi(x))$, $f(\theta u) = \theta f(u)$, and

$$R(x,r) = f(u^*)\cdot\pi(r)$$

where $f(u)$ is an index function of outputs and u^* yields $R(x,r)$.

The more or less common practice of price deflating dollar values to convert to physical quantities is thus seen to be strictly justified if and only if the production structure is homothetic.

7.6 Duality Between Cost (Revenue) Function and Distance Function

It is a remarkable fact that under the general axioms $(P.1,\ldots,P.6) \Leftrightarrow (L.1,\ldots,L.6)$ for production functions that cost function and distance function $\Psi(u,x)$ defining $u \to L(u)$ are dually related by

$$Q(u,p) = \underset{x}{\text{Min}} \{p\cdot x \mid \Psi(u,x) \geq 1\}$$

$$\Psi(u,x) \leq \underset{p}{\text{Inf}} \{p\cdot x \mid \phi(u,p) \geq 1\}$$

and the revenue (return) function and distance function $\Omega(x,u)$ defining $x \to P(x)$ are dually related by

$$R(x,r) = \underset{u}{\text{Max}} \{r\cdot u \mid \Omega(x,u) \leq 1\}$$

$$\Omega(x,u) \geq \underset{r}{\text{Sup}} \{r\cdot u \mid R(x,r) \leq 1\}$$

See [7.7] and [7.9] for proofs. If, in addition, inputs and outputs are disposable and the map sets $P(x)$ and $L(u)$ are convex, the equality sign holds in the second of each pair, implying that input structure $u \to L(u)$ and output structure $x \to P(x)$ are completely determined by cost and revenue (return) function respectively.

Our concern here is not so much whether and how cost

and revenue structure may be observed to impute input and output structure, but the imputation of shadow (accounting) prices, for which we need only the general axioms.

For given resource (input A) vector x^0 and price vector r^0 for output vectors, a shadow price vector p_s is determined for x^0 as a solution of the problem

$$\operatorname{Inf}_p \{p \cdot x^0 \mid Q(u^*,p) \geq R(x^0,r^0)\}$$

where u^* is an output vector yielding $R(x^0,r^0)$, and the shadow price vector p_s is value minimizing price for the given input vector x^0 which yields a value for x^0 which is at least as large as the maximal output value obtainable for x^0.

For given output vector u^0 and price vector p^0 for input vectors, a shadow price vector r_s is determined for u^0 as a solution of the problem

$$\operatorname{Sup}_r \{r \cdot u^0 \mid R(x^*,r) \leq Q(u^0,p^0)\}$$

where x^* is an input vector yielding $Q(u^0,p^0)$, and the shadow price vector r_s is a value maximizing price for the given output vector u^0 which yields a value for u^0 not exceeding the minimal cost of getting u^0 under prices p^0 for inputs.

For any feasible pair of vectors u^0 and x^0, that is, such that $u^0 \in P(x^0)$, the two problems

$$\operatorname{Inf}_p \{p \cdot x^0 \mid Q(u^0,p) \geq 1\}$$

$$\operatorname{Sup}_{r} \{r \cdot u^0 \mid R(x^0, r) \leq 1\}$$

determine shadow prices p_s and r_s simultaneously for input and output vectors such that the imputed value of u^0 by value maximizing price vector r_s does not exceed the imputed value of x^0 by value minimizing price vector p_s. Thus, both inputs and outputs may be shadow priced solely from the physical structure of production. In the case of linear structure of production these shadow prices are determined by linear programs.

7.7 Indirect Production Functions

As a final topic of this survey we consider indirect production functions as a construction for cost-return and cost-benefit analysis (see [7.8]).

A cost indirect production function arises when one seeks to find the output vectors that may be obtained by a cost rate expenditure $c > 0$ under prices p for inputs. The correspondence (indirect production function) is a relationship

$$\left(\frac{p}{c}\right) \to G\left(\frac{p}{c}\right) = \{u \mid Q(u,p) \leq c\}$$

The output sets $G(p/c)$ have properties similar to those of $P(x)$, except that $G(p/c)$ is not necessarily closed and bounded unless $p > 0$, and the sets $G(p/c)$ are not necessarily convex when $P(x)$ and $L(x)$ are convex unless the graph of $x \to P(x)$ $(u \to L(u))$ is convex, that is, nonincreasing returns to scale.

A return indirect production function is defined by

$$\left(\frac{r}{\mathbb{R}}\right) \to S\left(\frac{r}{\mathbb{R}}\right) = \{x \mid R(x,r) \geq \mathbb{R}\}$$

for positive return rate $\mathbb{R}$ and price vector r for outputs to relate return rate deflated price vector r to the set of input vectors which yield at least the return rate $\mathbb{R}$ under prices r for output vectors. Here the input sets $S(r/\mathbb{R})$ have properties like those

of the input sets $L(u)$, except that $S(r/\mathbb{R})$ is convex if the graph of $u \to L(u)$ is convex, that is, $P(x)$ $L(u)$ convex $\not\Leftrightarrow$ $S(r/\mathbb{R})$ convex .

In case the direct production structure $u \to L(u)$ is homothetic, the output sets $G(p/c)$ of the production function $(p/c) \to G(p/c)$ are defined by

$$G\left(\frac{p}{c}\right) = \left\{ u \mid f(u) \leq F\left(\frac{c}{P(p)}\right) \right\}$$

and the isoquants of $(p/c) \to G(p/c)$ are defined simply by the equation

$$f(u) = F\left(\frac{c}{P(p)}\right)$$

Similarly, when the direct production structure $x \to P(x)$ is homothetic;

$$S\left(\frac{r}{\mathbb{R}}\right) = \left\{ x \mid \phi(x) \geq F^{-1}\left(\frac{\mathbb{R}}{\pi(r)}\right) \right\}$$

and the isoquants of the production function $(r/\mathbb{R}) \to S(r/\mathbb{R})$ are defined by the equation

$$\phi(x) = F^{-1}\left(\frac{\mathbb{R}}{\pi(r)}\right)$$

If both $x \to P(x)$ and $u \to L(u)$ are homothetic

$$\left(\frac{p}{c}\right) \to G\left(\frac{p}{c}\right) = F\left(\frac{c}{P(p)}\right) \cdot P_f(1)$$

$$\left(\frac{r}{\mathbb{R}}\right) \to S\left(\frac{r}{\mathbb{R}}\right) = F^{-1}\left(\frac{\mathbb{R}}{\pi(r)}\right) \cdot L_\phi(1)$$

and the map sets for the indirect production functions are developable in terms of the sets $P_f(1)$ and $L_\phi(1)$ for the direct structure, by the same scaling functions as in the direct case with arguments in terms of price deflated cost and return rate.

If global prices are assumed for inputs and outputs, cost-return relationships are developed simply in terms of the following two functions

$$\mathbb{R}\left(\frac{p}{c},r\right) = \operatorname*{Sup}_{u}\left\{ r\cdot u \mid u \in G\left(\frac{p}{c}\right)\right\}$$

$$K\left(\frac{r}{\mathbb{R}},p\right) = \operatorname*{Inf}_{x}\left\{ p\cdot x \mid x \in S\left(\frac{r}{\mathbb{R}}\right)\right\}$$

expressing the maximal return obtainable with a cost rate c under prices p and r for inputs and outputs, and the minimal cost of getting a return rate $\mathbb{R}$ under prices p and r.

One expects these two functions to be reciprocally related under suitable conditions. If both direct production functions are homothetic

$$\left(\frac{\mathbb{R}\left(\frac{p}{k},r\right)}{\pi(r)}\right) = F\left(\frac{K\left(\frac{r}{\mathbb{R}},p\right)}{P(p)}\right)$$

that is, real return rate and real cost rate are inverse functions of each other. This equation has significance for econometric studies. If one assumes that for each price vector pair p and r production is carried out so that revenue and cost are maximally and minimally related, price deflated return rate and price deflated cost rate may be plotted against each other to estimate the function $F(\cdot)$ for the joint production function

$$f(u) = F(\phi(x))$$

where $f(u)$ and $\phi(x)$ are index functions for outputs and inputs.

There are obvious difficulties in treating p and r as globally applicable price vectors for determining value. If a social utility function $V(u)$ is postulated for output vectors, one may define an indirect production function by

$$E \to \sum(E) = \{x \mid x \in L(u)\ ,\ V(u) \geq E\}$$

where $E \varepsilon (-\infty,+\infty)$ is a real number, and $V(0) = 0$ with $V(u) < 0$ implying u is less preferred to the null output vector. It is allowed that some of the components of u are unwanted. A map set $\Sigma(E)$ defines those input vectors which yield output vectors u at least as preferred as those of the indifference class to which E is assigned. The sets $\Sigma(E)$ have properties like those of the input sets $L(u)$.

Relative to the production function $E \to \Sigma(E)$, one may define the minimal cost of achieving output vectors at least as preferred as those of the indifference class to which E is assigned, by

$$X(E,p) = \underset{x}{\text{Inf}} \{p \cdot x \mid x \varepsilon \sum(E)\}$$

which may serve to cardinalize the ordinal utility function $V(u)$. For each value E of $V(u)$, take $X(E,p)$ as the corresponding cardinal value. Then under a supposed global price p for inputs, a cost-benefit relationship is defined by

$$X\left(V^*\left(\frac{p}{c}\right),p\right)$$

where

$$V^*\left(\frac{p}{c}\right) = \text{Sup}\left\{ V(u) \mid u \varepsilon G\left(\frac{p}{c}\right)\right\}$$

which in resource terms may provide cardinal measure of the maximal benefits obtainable at a cost rate $c > 0$ under prices p for inputs.

7.8 Postscript for Dynamic Models

One area of interest for the Navy, concerning production functions, is the construction of ships. Here the steady state net output model of production fails to be useful. If one observes a shipyard there are many production activities being carried out each day, yet no output emerges. Months pass, yes, even years pass, and still no net output. Finally a ship emerges. A model for net outputs does not address itself to ship construction in sufficient detail. The daily production

consists of a great variety of intermediate products, which must be explicit in some way for a suitable production model. On closer examination, the various daily production tasks are not daily repetitive, that is, they need to be time dated, and there is a complex relationship between tasks as to order and precedence representable by a linear graph like that used in PERT and critical path scheduling. Thus, one is deeply involved in a dynamic structure for ship construction.

How, then, may one define the production function? Let the arcs of a linear graph denote production activities as in the linear, steady state model with intermediate products. For each activity, the outputs are intermediate products and require as inputs exogenous factors as well as intermediate products, with input rates proportional to activity intensity. In some way, one may compute on the graph the minimal time from start to finish for construction, under related time patterns of exogenous inputs. Then the production function may be defined as a functional (in scalar output case)

$$F\begin{pmatrix} t \\ x(\tau) \\ t_0 \end{pmatrix} = \frac{1}{T}$$

where T is the minimal time on the graph (the so-called period of production), and $x(\tau)$ is a vector of functions $x_i(\tau)$ each of which is defined on some time interval $[t_0,t]$ to specify the time rates at which exogenous inputs are available over time, being zero outside this interval. The functional

$$F\begin{pmatrix} t \\ x(\tau) \\ t_0 \end{pmatrix}$$

defines the instantaneous output rate dynamically as a functional of time histories of availability of exogenous inputs.

Little is known about this kind of production function,

although it was suggested by Evans [7.1] in 1930, and research must be undertaken to develop the notion for production activities that constitute a major construction of a single net output. A generalization of the theory of production functions along such lines is important for economic theory as well as having obvious advantages for naval applications.

References

[7.1] Evans, G. C. (1930). Mathematical Introduction to Economics. McGraw-Hill. Appendix II.

[7.2] Gale, D. (1960). The Theory of Linear Economic Models. McGraw-Hill.

[7.3] Johansen, L. (1972). Production Functions. North Holland.

[7.4] Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. in T. C. Koopmans (ed.) Activity Analysis of Production and Allocation. Cowles Commission Monograph No. 13. Wiley. 33-97.

[7.5] Leontief, W. W. (1941). The Structure of the American Economy, 1919-1929. Oxford University Press.

[7.6] Schumpeter, J. A. (1966). History of Economic Analysis. Oxford University Press. 1031.

[7.7] Shephard, R. W. (1970). Theory of Cost and Production Functions. Princeton University Press.

[7.8] Shephard, R. W. (1974a). Indirect Production Functions. Mathematical Systems in Economics, No. 10, Verlag Anton Hain, Meisenheim am Glan.

[7.9] Shephard, R. W. (1974b). Semi-homogeneous production functions and scaling of production. Production Theory. in W. Eichhorn et al. (eds.) Lecture Notes in Economics and Mathematical Systems. Springer. 253-285.

[7.10] Shephard, R. W., and R. Färe (1974). The law of diminishing returns. Zeitschrift für Nationalökonomie 34 69-90.

Chapter 8

BRANCH-AND-BOUND REVISITED: A SURVEY OF BASIC CONCEPTS AND THEIR APPLICATIONS IN SCHEDULING*

S. E. Elmaghraby and A. N. Elshafei**
North Carolina State University

8.1 Preliminaries

The term "branch-and-bound" (B&B) has increasingly become a household term among students and researchers in the field of scheduling and sequencing. In this chapter we shall take a fresh look at this approach and assess its content, utility, and potential. In delineating the subject matter of our discussion, perhaps it is equally valid to emphasize that which is not among our aims. This chapter is not a comprehensive survey of B&B concepts and applications. Several survey articles that have appeared in recent years serve that function adequately, if not superbly; see, for example, References [8.1, .4, .21, .36, .40]. Nor does this paper aspire to be a comparative evaluation of the very many B&B approaches that have been proposed in the open literature to solve one scheduling problem or another. For examples of such studies, the reader is referred to the papers of Ashour [8.2], Ashour and Quraishi [8.3], Davis [8.8], and Kan [8.34], among others.

What we do wish to present is an inventory of the basic concepts underlying the "theory" of B&B; we wish in fact to establish that such theory exists and to illustrate these basic concepts by examples from the field of scheduling and sequencing. In this we are

---

*The preparation of this chapter was partially supported by the Office of Naval Research under Contract N00014-70-A-0120-0002, by the National Science Foundation under Grant P1K1470-000, and by the Army Research Office-Durham under Contract DA-ARO-D-31-124-72-G106, with North Carolina State University.

**Now at the Institute of National Planning, Cairo, Egypt

motivated by two objectives. The first is to summarize, in what we hope is a convenient place, the multitude of concepts that have emerged over the past few years. We hope that such a summary will provide a handy reference and basic understanding to student and researcher alike. The second is to help the profession assess the current and future potential of this approach. In this respect, one may compare B&B as a problem solving approach, to simulation which is another, and by now a very popular problem-solving approach. One may then ask fundamental questions similar, but not necessarily identical, to those asked in the study of simulation. For instance, in Monte Carlo simulation one often raises the question of variance-minimizing techniques. In B&B one may ask questions relative to the rate of convergence to the optimum.

As much as possible we shall draw our examples from the field of scheduling and sequencing. However, since problems of scheduling (and sequencing) are almost universally modeled as integer or mixed programming problems (linear or nonlinear), we shall feel free to illustrate some concepts with reference only to the integer (or mixed) program, without the need to motivate the model by the scenario of the scheduling problem. Ordinarily, we shall be dealing with integer linear problems (ILP) and, in particular, with 0,1 ILPs. As is well known, an ILP can be translated into a 0,1 ILP by the simple binary expansion of the variables. In a couple of instances, we could not find examples from scheduling and, to the best of our knowledge, none exist that use a particular concept. Then we took the liberty to illustrate by examples from other fields of application, such as location-allocation. We do not feel particularly apologetic about taking such liberty since these problems are themselves modeled as integer (linear or nonlinear) programs. Such models provide the link to problems in scheduling.

In the sequel we shall be talking about "partial solutions" and "completions." The term "partial solution" is actually a misnomer, since it refers to something that provides no solution whatsoever to the original problem. For instance, a schedule of a subset of the jobs, or a series of cities visited by the salesman in the traveling salesman problem (TSP), are

referred to as partial solutions, yet they provide no "solution" to the problems posed, which are: a complete schedule of all jobs in the first case, and a complete tour over all cities in the second case. The reader will hopefully bear with this misuse of language. By the completion of a partial solution we mean the specification of the values of the remaining variables so that their union with the partial solution yields a point in the original solution space. A partial solution is said to be fathomed if one of the two following conditions is satisfied.

(1) We determine that its best feasible completion is better (yields a better objective value) than the best feasible solution known to date (assuming one is in hand).

(2) We determine that the partial solution has no feasible completion better than the incumbent (this includes infeasibility, which is translated into infinite penalty).

The concept of fathoming is illustrated in Example 8.3.

8.2 Fundamentals

The approach of B&B is basically a heuristic tree search in which the space of feasible solutions, which may contain a very large (or denumerable) number of points, is systematically searched for the optimum. According to Mitten and Warburton [8.42], "the search proceeds iteratively by alternately applying two operations: subset formation and subset elimination. In the former, new subsets of alternatives are formed, while in the latter some subsets of alternatives may be eliminated from further consideration. The procedure terminates when a collection recognized to contain only optimal solutions is reached." The search has two guiding principles: first, that every point in the space is enumerated either explicitly or implicitly, and, second, that the minimum number of points be explicitly enumerated. (We view B&B as an approach for implicit enumeration, though we concede that, mainly due to historical coincidences, the label "implicit enumeration" has been applied to approaches that need

not employ the "bounding" feature of B&B.)

The implicit enumeration of feasible points is accomplished through dominance (which may or may not employ bounding) and feasibility considerations. Each of these concepts will be discussed in greater detail below, but first we give a laconic description of them to afford the uninitiated reader a general grasp of the subject. The basic idea in B&B is to divide the feasible space, denoted by S, into subsets $S_1, S_2, \ldots, S_k$ which may or may not be mutually disjoint. Assuming that the optimum falls in subset S_k, a bound on its value is determined: an upper bound (u.b.) in the case of maximization, and a lower bound (ℓ.b.) in the case of minimization or, better still, both an upper and a lower bound in either case. Based on such bounds two actions may take place: (i) a particular subspace S_k is selected for more intensive search by further partitioning into its subsets (this is the branching, or "formation" function); (ii) some feasible points (subspaces) are declared "noncandidates" for the optimum, and thus are eliminated from further considerations. This latter idea is one of "dominance" since it is based on the determination that any element of a particular subset S_i is better (or worse), in the sense of the criterion function, than any element in another subset S_j. Then indeed we may declare the points in S_j (or in S_i) as noncandidates for the optimum and eliminate them from further analysis.

While dominance may be established on the basis of the bounds evaluated on subsets S_k, it is also true that dominance can be established independent of any bounding considerations. In some circles (especially in the scheduling literature) these are referred to as "elimination" procedures. The final result is the same, namely, it establishes that certain subspaces cannot contain the optimum because they are dominated by other subsets. A similar idea lies behind the feasibility considerations. They arise because in the majority of cases one is forced to hypothesize a rather "rich" original space S. At some stage of analysis,

if it can be established that certain subsets of S are in fact infeasible (in the sense of violating some constraint of the problem), then indeed such subsets can also be eliminated from further study.

Heuristics enter the tree search in all three basic phases of the approach: in the definition of the partitioning procedure, in the calculation of the bounds, and in the philosophy of searching the tree. But we wish to draw the reader's attention to the following important and rather crucial distinction: the formal structure of B&B admits the use of heuristics (as does the simplex algorithm of linear programming). However, these are "reliable heuristics" in the sense that if they run to completion, the optimum will be achieved. Furthermore, if the procedure is terminated before it has achieved the optimum, it yields a bound on the error committed. (This is in sharp contrast to "heuristic problem-solving procedures" which lay no claim to either optimality or to measuring the error committed at premature abortion.)

A more formal definition of the B&B procedure was advanced by Mitten [8.39] in 1970 which was expanded upon in later work in 1973 by Mitten [8.40], and Mitten and Warburton [8.42]. Mitten defines the operations of "branching," "bounding," and "branch-and-bounding" in terms of set functions. The necessary properties of each function were given in terms of operator and operands, which map all the known concepts of B&B into topological domains. He establishes the relations between the B&B recursive function and the set of optimal feasible solutions by postulating various analytic and topological conditions such as continuity, completeness, and compactness. In the case of finite solution space, the convergence of the B&B recursive function is easily seen. However, in denumerable or nondenumerable spaces, Mitten demonstrates that if the B&B recursive function is a contraction mapping in a complete metric space with appropriately defined elements, then fixed-point theorems could be invoked to establish convergence.

To gain more insight into Mitten's construction, we assume that it is desired to solve the problem: maximize $f(x)$ for $x \in X$. (For example, X may be the integer feasible points in an ILP.) Typically, B&B

proceeds by searching the space $T \supseteq X$ for the set $\sigma^* \triangleq \{x \in X: f(x) = f^*\}$ of optimal solutions, where $f^* = \sup_{x \in X} f(x)$ and where $\triangleq$ means "is defined to equal." The search proceeds by examining subsets $\sigma \in X$, and collections of such subsets. Let S denote the family of all possible collections of subsets that could be encountered by a given B&B procedure. As shorthand notation, let $\cup(s)$ denote a subset of X comprised of the elements in $\bigcup_{\sigma \in s} \{\sigma\}$; that is

$$\cup(s) \triangleq \bigcup_{\sigma \in s} \{\sigma\} \quad \text{where} \quad s \in S .$$

As mentioned above, alternative possibilities in B&B are considered in sets rather than one at a time. Furthermore, B&B examines successively smaller and smaller subsets of X (the subset formation operation), always eliminating those subsets that can be shown not to contain an optimal solution (the elimination operation). It is assumed that once sets are "small enough" in some sense, then there is a procedure available for distinguishing the optimal solutions from the nonoptimal solutions, the so-called fathoming procedure. Therefore, let S^- denote the set of <u>fathomable</u> collections $\{s\}$; here s is a fathomable collection iff $\sigma \in s$ satisfies $\sigma \subset \sigma^*$ or $\sigma \cap \sigma^* = \phi$. We assume that a procedure is available for separating one from the other. That is, $s \in S^-$ iff the following hold.

(a) $s = s_1 \cup s_2$ with $\sigma \subseteq \sigma^*$ for every $\sigma \in s_1$ and $\sigma \cap \sigma^* = \phi$ for every $\sigma \in s_2$

(b) There is a means available for forming the collections s_1 and s_2.

As a minimum requirement, we insist that any collection of singleton sets (sets containing one point of X each) is fathomable, since such sets cannot be subdivided. One may now state the objectives as: find a collection $s^* \in S$ such that $\cup(s^*) \subseteq \sigma^*$ and

$\cup(s^*) = \phi$ only if $\sigma^* = \phi$. Mitten defines the B&B procedure in terms of the set operations called branching, upper bounding, lower bounding, and, finally, branch-and-bounding. Let S_F (for formation) and S_E (for elimination) be two subfamilies of S. Branching may be defined as a function $F: S_F \to S_E$ such that for each $s \in S_F$ the following hold.

(a) $F(s) = \bigcup_{\sigma \in s} \{d(\sigma)\}$, where $d(\sigma)$ is either σ or a collection of proper subsets of σ whose union is σ

(b) $F(s) = s$ iff $s \in S^-$

In words, this latter condition (a) states that each σ in s either remains unchanged under $F(s)$ or is broken up into a collection of proper subsets. This is illustrated in Figure 8.1, in which $s = \{\sigma_1, \sigma_2, \sigma_3\}$; $d(\sigma_2) = \sigma_2$; $d(\sigma_3) = \sigma_3$, but σ_1 has been "broken up" into four (disjoint) subsets. Clearly, $\bigcup_j \sigma_{1j} = \sigma_1$.

Upper bounding is a real-valued function $u: \cup(S_E) \to \underline{R}$ with the following properties.

(a) $u(\sigma) \geq f(x)$ for all $x \in \sigma \in \cup(S_E)$

(b) $u(\sigma) \geq u(\sigma_0)$ if $\sigma_0 \subseteq \sigma$; $\sigma_0, \sigma \in \cup(S_E)$

(c) $u(\{x\}) = f(x)$, $x \in \sigma$

These latter conditions (a) and (b) follow from the common concepts of upper bounds and set inclusion. Condition (c) ensures that the upper bound on singleton subsets is the "value" of that point under the mapping f. Lower bounding is a real-valued function $\ell: S_E \to \underline{R}$ such that the following hold for any $s \in S_E$.

(a) $\ell(s) \leq f^*$

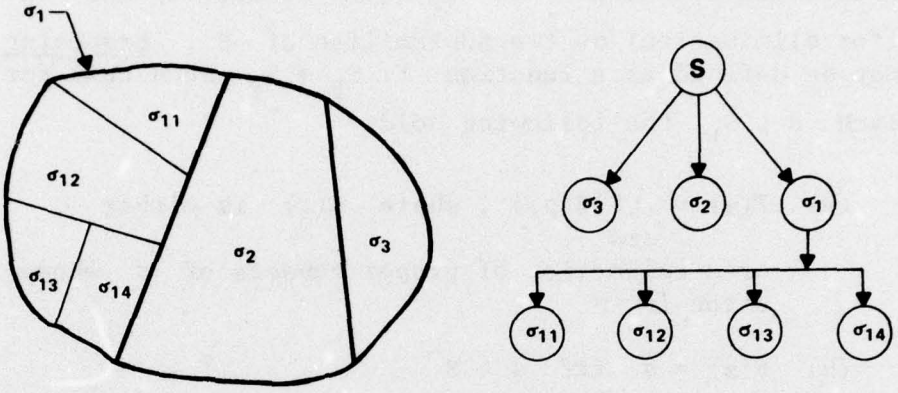


Figure 8.1 - Partitioning and branching.

(b) $\ell(s) \leq \ell(F(s))$

(c) $\ell(s) \geq f(x)$ for any $x \in s$

(d) If $s' \subset s$ is such that, for every $\sigma' \in s'$ either $u(\sigma') = -\infty$ or $u(\sigma') < \ell(s)$, then $\ell(s') = \ell(s)$.

These latter conditions (a) and (b) follow from the common concepts of lower bound and set division into subsets. Condition (c) ensures that the lower bound of a singleton subset is tight. Condition (d) guarantees that infeasible sets $(u(\sigma') = -\infty)$ or dominated sets $(u(\sigma') < \ell(s))$ cannot affect the value of the lower bound $\ell(s)$.

In Mitten's view, the bounding operation is an elimination operation through infeasibility and dominance. He defines it as a function $E: S_E \rightarrow S_F$ defined for $s \in S_E$ by

$E(s) = s - \{\sigma \epsilon s: u(\sigma) = -\infty \text{ or } u(\sigma) < \ell(s)\}$

The strict inequality in the above statement implies strong bounding, since all optimal solutions in σ^* are retained. If an inequality is substituted, the resulting bounding operation is said to be weak since we then ensure that at least one element of σ^* will be retained. Finally, the B&B recursive operation is a function $G: S_F \to S_F$ defined by $G(s) = E(F(s))$. In other words, the successive formation and elimination of subsets is the heart of the procedure, hopefully leading to an optimum without the need to enumerate all singleton subsets.

If X contains finitely many points, it can be shown that $s^n = G(s^{n-1})$, for some finite $n > 1$, is an element of the fathomable set S^- , so that the procedure will terminate in a finite number of iterations. In the case X is not finite, Mitten shows that G will not "cycle" provided that each collection in S_F and S_E contains only finitely many sets. (Cycling means that there exists an m such that $G^m(s) = s$ and $s \epsilon S_F - S^-$.) Note that even though a procedure may never cycle, it may not terminate in a finite number of steps. With this formal structure established, Mitten proceeds to illustrate his concepts by two examples: ILP and sequential unimodal search. This latter illustration is interesting since it claims to be the following.

(i) An example in which neither the procedure nor the sets involved are finite.

(ii) The only currently known application of B&B methods employing a branching rule that can be demonstrated to be optimal (the Fibonacci search).

This led Mitten to the following two conclusions. First, that the existence of an optimal branching rule

for the sequential unimodal search suggests some interesting avenues of investigation in other areas of application. Second, that since attempts to extend the sequential search method to higher dimensions $\underline{\underline{R}}^n (n>1)$ have been notably unsuccessful, perhaps a fresh attack on the problem in $\underline{\underline{R}}^n$ via B&B would provide a new perspective. We wish only to remark that viewing sequential unimodal search as an application of B&B may raise some eyebrows, since none of the concepts usually associated with B&B are present in the standard search pattern, including the optimal pattern. The viewing is justified, however, if one sticks to the formal definition of B&B's search as composed of set formation and set elimination, both of which are indeed present in sequential unimodal search.

8.3 Branching

Branching proceeds by dividing the solution space into subspaces, which are themselves divided into subspaces, and so forth, until subsets containing exactly one point each are reached. The graphic representation is a tree, the search tree, whose numbering runs opposite to the set content. Thus, S_0 is the empty set ϕ which represents in fact the whole space S before any division has taken place. A terminal node of the tree, S_M, M large, contains a complete solution X which represents in fact a singleton set. Intermediate nodes of the search tree generally represent partial solutions generically represented by S_k. Hereafter we use the terms "branching" and "dividing the solution space" synonymously. The choice of the node from which to branch is basically a decision related to the philosophy of searching the tree, which is open to the use of heuristics.

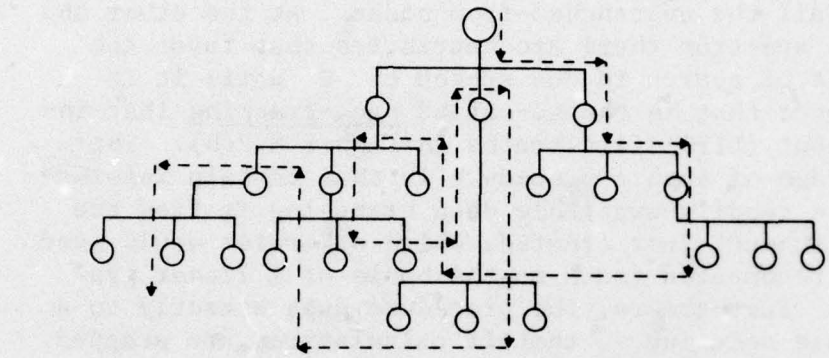
Basically, there are two extreme philosophies with innumerable intermediate variations. On the one end of the spectrum there are heuristics (for example, branch from the node with the smallest lower bound) that favor the nodes higher up the tree. In this case, the construction of the search tree will proceed "horizontally"; this is the so-called jump-tracking (or "flooding")

illustrated in Figure 8.2(a). The advantage of this procedure is its economy in the number of singleton sets created prior to the determination of the optimum. Its disadvantage lies in the vast memeory required to store all the unbranched-from nodes. At the other end of the spectrum there are heuristics that favor the pursuit of search in one subset of S until it is fathomed; that is the so-called back-tracking last-in-first-out (LIFO) illustrated in Figure 8.2(b). The advantage of such a procedure is that certain information is readily available when branching is from the node (subset) just created, which otherwise would need to be recomputed (such as the basis of a linear program). Furthermore, the procedure goes directly to a feasible solution so that if calculations are stopped before optimality is achieved, there is available a feasible solution as well as an upper (or lower) bound on the optimum value.

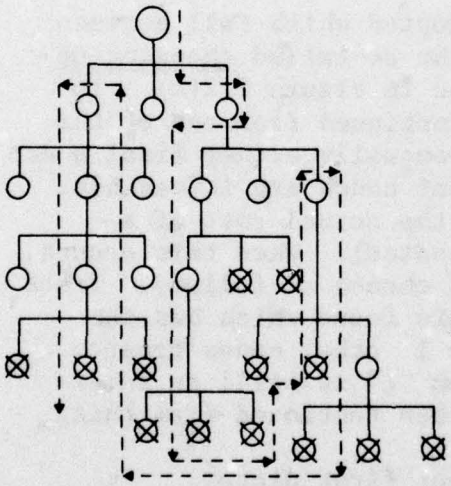
Other procedures may be adopted which fall between these two extremes, such as the so-called choosing up the tree procedure illustrated in Figure 8.2(c). In this case, branching always continues from one of the m nodes just created until eventually either final nodes are obtained, or all descendant nodes are infeasible, or their lower bounds exceed the actual cost of a known solution (they are dominated). When this occurs, the next intermediate node is chosen as follows. Track up the tree until a node ξ is found which has the property that one of the $m - 1$ other nodes created when branching took place from ξ is still an intermediate node. Branching is then continued from this intermediate node.

We are now able to state our first dictum.

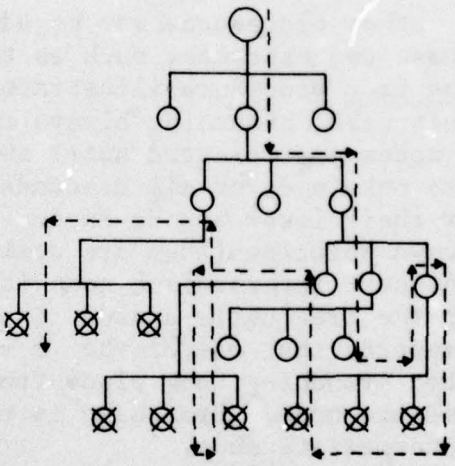
I. The Branches Need Not Be a Partition. The definition of the partitioning procedure is synonymous with the definition of the branching procedure. In the majority of cases there are several ways by which S may be partitioned: the decision is basically a heuristic one. Different procedures lead to different numbers of subsets of any subspace S_k and, consequently, different numbers of "stages" or "levels" of the search tree. The concept we wish to advance here as Dictum I is that the division of a subspace S_k



(a) Jump-tracking



(b) Back-tracking
⊠ Fathomed nodes



(c) Choosing-up-the-tree
⊠ Fathomed nodes

Figure 8.2 - Three patterns of search in branching.

into subsets $S_{k1}, S_{k2}, \ldots, S_{kr}$ need not be a partition in the sense that $S_{ki} \cap S_{kj} = \phi$. This is true because the subsets may have some points in common. We illustrate this concept with two examples.

Example 8.1. Consider the problem of scheduling N jobs on M_t identical machines available in period t (say day t), $t = 1,2,\ldots,H$, where H is the "planning horizon." The jobs are related by precedence constraints. A job j has: processing time y_j during which it occupies one machine uninterrupted; a desired completion time (the so-called "due date") d_j; and a cost $c_j(T_j - d_j)$ which is a function of the difference between the actual completion time of the job T_j and its due date. (The function c_j is quite general except for the mild restriction that it be nondecreasing away from d_j.) It is desired to find the schedule of the N jobs with minimum total cost.

An ILP model was advanced by Elmaghraby [8.13]. Recognizing the computational difficulties entailed in a frontal attack, he proposed a B&B procedure in which the "levels" of the search tree correspond to the jobs, and a subspace S_k defines a partial solution in which the first $k - 1$ jobs have specified start times. Notationally, this is given by $\lambda_{is_i} = 1$ for $i = 1,2,\ldots,k-1$ in which λ_{it} is a 0,1 variable denoting the start or non-start of job i in period t and s_i represents the start time of i. Because of the precedence constraints, let a_k denote the earliest availability of job k (its earliest start time), and b_k its latest completion time (an "absolute" deadline beyond which the job may not be completed). Clearly, $\lambda_{ks_k} = 1$ for some value of s_k in the interval $a_k \leq s_k \leq b_k - y_k + 1$. Hence the subspace S_k is partitioned into $b_k - a_k - y_k + 1$ subsets, which may

be represented by branches in the search tree as shown in Figure 8.3. Note that some of these subsets may be infeasible due to the machine availability constraints, in which case the subsets may be eliminated from further considerations.

An alternative approach may run as follows. Let the levels of the search tree correspond to time periods. In any period t, a number of jobs, say $n \leq N$, are eligible for being started (by virtue of the precedence relations). The machine availability constraints may limit the number of jobs that can be started simultaneously to (the binomial coefficient) $Q = C(n, M_t - r)$ alternatives, where $r < M_t$ is the number of machines "committed" in period t as a consequence of the partial scheduling of the first $k - 1$ jobs. Suppose we enumerate all such feasible "bunches" of activities, and denote them by $S_{t,1}, S_{t,2}, \ldots, S_{t,Q}$. Then we may branch to $Q + 1$ subsets corresponding to the Q ways in which activities may be initiated at time t, plus the state in which none are initiated.

Comparing the two procedures, it is evident that in the latter the various subsets need not be mutually

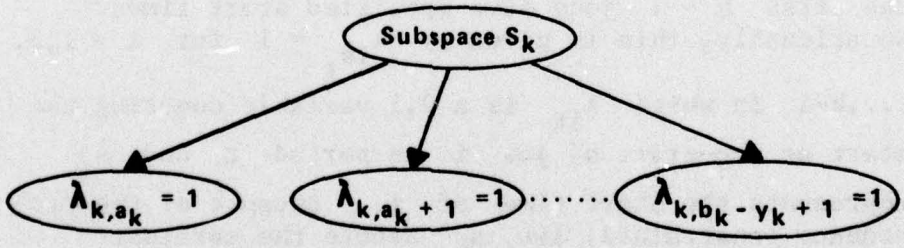


Figure 8.3 - The partitioning of S_k.

exclusive and hence the subdivision of S_k is not a partition. Moreover, the first procedure leads to a tree of N levels, while the second to a tree of H levels.

Example 8.2. Consider the well-known traveling salesman problem (TSP) over N cities, modeled by the following ILP.

$$\text{minimize} \quad \sum_i \sum_j c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_j x_{ij} = 1 \quad \text{for all } i \,, \quad \sum_i x_{ij} = 1 \quad \text{for all } j$$

$$u_i - u_j + N x_{ij} \leq N - 1 \,, \quad i = 2,\ldots,N \,,$$

$$j = 2,\ldots,N \,, \quad i \neq j$$

$$x_{ij} = 0,1 \,, \quad u_i \geq 0 \quad \text{and integer for all } i$$

We propose three modes of branching, two of which partition the subspaces but the third does not. Let S_k define a path from the home city 1 to city i_k. A tour T is any permutation of the $N - 1$ non-home cities and a tour T_k is a tour that includes path S_k. We also call S_k a subtour. Then arc $(i_k j)$ either belongs to tour T_k or not. The three modes of branching are as follows.

(i) The path S_k generates two subsets of the tours T_k for each city $j \notin \{1, i_1, i_2, \ldots, i_k\}$: the tours wherein S_k is extended by continuation on arc $(i_k j)$ which we denote by S^1_{kj} ; and the tours denoted by S^2_{kj} wherein S_k appears but arc $(i_k j)$ does

not. Figure 8.4(a) illustrates this mode of branching which is basically the branching rule proposed by Little et al. [8.37]. The branch S^1_{kj} eliminates from consideration (due to infeasibility) the row i_k in the distance matrix and any other entry (ji_r) that would form a subtour with the current partial schedule $(i_r \varepsilon \{1, i_1, \ldots, i_k\})$. On the other hand, the branch S^2_{kj} simply eliminates the entry $(i_k j)$ in the distance matrix. Recalling the bounding method of [8.37], it is obvious that the positive assertion of S^1_{kj} is more potent than the negative assertion of S^2_{kj}.

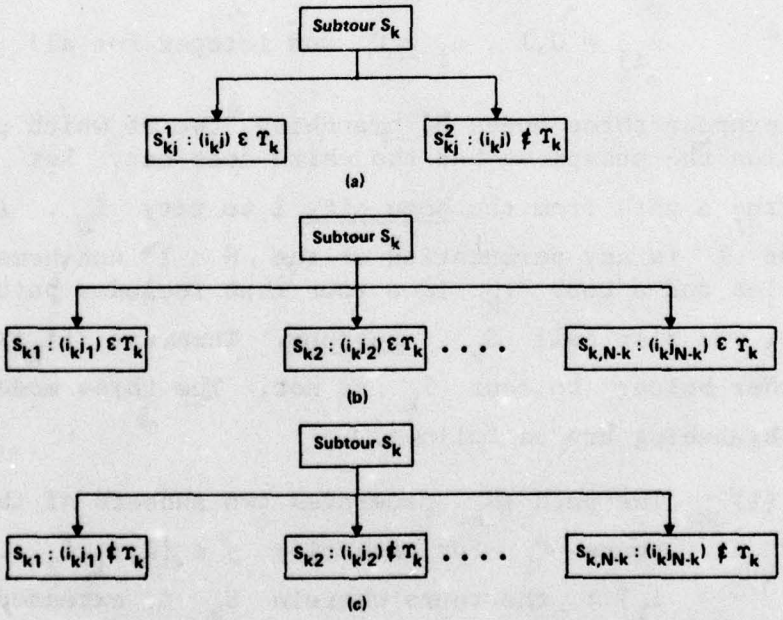


Figure 8.4 - Three modes of branching.

(ii) Still assuming that S_k defines a path to city i_k, specify the subsets according to the next arc included in the tour T_k. Thus $s_{k1} = S_k \cap (i_k j_1)$; $S_{k2} = S_k \cap (i_k j_2), \ldots,$ $S_{k,N-k} = S_k \cap (i_k j_{N-k})$, and there are as many branches from S_k as there are cities still to be visited. This mode is illustrated in Figure 8.4(b).

(iii) Specify the subsets of S_k by the arcs not in the tour. Then we write S_{k1} to denote the set of all tours T_k in which the arc $(i_k j_1)$ does not appear; S_{k2} denotes the set of all tours T_k in which the arc $(i_k j_2)$ does not appear; and so on. Clearly this is not a partition of the tours T_k since, for instance, both subsets S_{k1} and S_{k2} contain all completions of the subtour S_k which contain neither city j_1 nor city j_2. Figure 8.4(c) illustrates this mode.

II. The Desirability of Nonredundancy of Completions. Undoubtedly, a desiratum would be that the branching scheme generates a sequence of nonredundant partial solutions; that is, that no completion of a partial solution in the sequence ever duplicates a completion of a previous partial solution that was fathomed. To heed this second dictum, it is obviously necessary and sufficient to have in all future subsets S_v, $v > k$, at least one element "complementary" to one in S_k. This, in turn, is indeed satisfied if we store S_k and generate the new subset S_{k+1} to be exactly S_k but with its last element the complement of the last element of S_k, and indicate in some fashion that S_k was

fathomed. (The storing of S_k is to comply with the requirement of (implicitly) enumerating all points in the solution space.) Compliance with this dictum is extremely difficult, and is rarely accomplished except in those instances where "complementarity" is obvious, such as 0,1 ILPs. The alternative to modeling the problem as a 0,1 ILP is to store all subsets in the tree (not just the unbranched-from subsets) and compare them to each newly-generated subset to eliminate duplication. (This is of fundamental importance in dynamic programming. In some sense, B&B relaxes this requirement, and the price paid for such relaxation is the possibility of duplication.) Our example to illustrate this concept is indeed taken from solutions to 0,1 ILPs.

Example 8.3. In the Geoffrion-Glover Approach to 0,1 ILP [8.26, .29], the problem is stated as follows for $c > 0$.

minimize cX

$$\text{s.t.} \quad AX + b \geq 0 \; ; \; x_j = 0,1 \; .$$

An S_k is a set of variables whose values have been assigned as either 0 or 1. The remaining variables are called "free variables." Suppose $S_k = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, and S_k is fathomed. Then create $S_{k+1} = \{x_{i_1}, x_{i_2}, \ldots, x_{i_{k-1}}, \underline{\chi_{i_k}}\}$ where $\chi_{i_k} = 1 - x_{i_k}$, and the underlining of χ_{i_k} is simply a visual indication that S_k has been fathomed. The following illustrates the two concepts of fathoming and nonredundant completion. The general logic may be shown schematically as in Figure 8.5. The application of this logic to the ILP proceeds as follows where the step number refers to the box number in Figure 8.5.

1. The most effortless completion of S_k , ignoring feasibility, is to put $x_j = 0$ for all the free

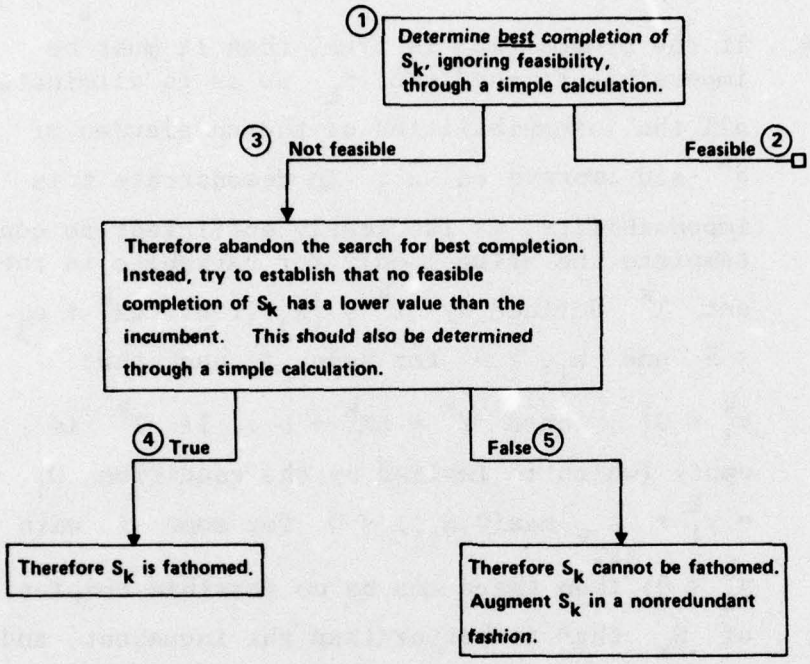


Figure 8.5 - General logic for fathoming.

variables. (Recall that we assumed all $c_j > 0$.)

2. Clearly, if the resultant solution is feasible, then indeed it is the best completion of S_k. Its value is easily determined, say z_k. If $z_k \leq \bar{z}$ (the value of the incumbent, assuming one exists; otherwise $\bar{z} = \infty$), then we adopt the current feasible solution as the best, and put $\bar{z} = z_k$. Otherwise, $z_k > \bar{z}$ and, *a fortiori*, no feasible completion of S_k has a lower value (of the objective function) than the incumbent; hence S_k is dominated, and again it is fathomed.

3. If the completion is not feasible, then instead of seeking the best completion of S_k (which may require extensive calculations) we try to establish the proposition stated in node 3.

4. If the proposition is true, then it must be impossible to complete S_k so as to eliminate all the infeasibilities of the completion of S^k and improve on $\bar{z}$. To demonstrate this impossibility, it is clearly sufficient to contemplate the value 1 only for variables in the set T^k defined by $T^k \triangleq \{x_j \text{ free: } cX^k + c_j < \bar{z} \text{ and } a_{ij} > 0 \text{ for some } i \text{ such that } y_i^k < 0\}$, where $Y^k = AX^k + b$. If T^k is empty (which is implied by the condition $D_i = y_i^k + \sum_{j \in T^k} \max(0, a_{ij}) < 0$ for some i with $y_i^k < 0$) then there can be no feasible completion of S_k that is better than the incumbent, and S_k is fathomed.

5. If the proposition is false, then S_k cannot be (easily) fathomed. Then the partial solution S_k must be augmented. Here, heuristics are adopted such as: add the variable that reduces total infeasibility the most; or add the variable that reduces infeasibility in the most number of constraints; or add the variable with the smallest $c_j > 0$ that still reduces infeasibility of at least one constraint; and so on. The application of these concepts to an ILP is given in the tree diagram contained in Figure 8.6. There the variables are denoted by their numbers; writing j as, for example in $S_3 = \{3, -\underline{2}\}$ means that $x_j = 1$ while $-j$ means $x_j = 0$; the underlining of a variable means that its complement was fathomed. Notice finally that the first of the stated procedures in 5. is used in Figure 8.6.

III. Alternate Criteria for Branching. In relating

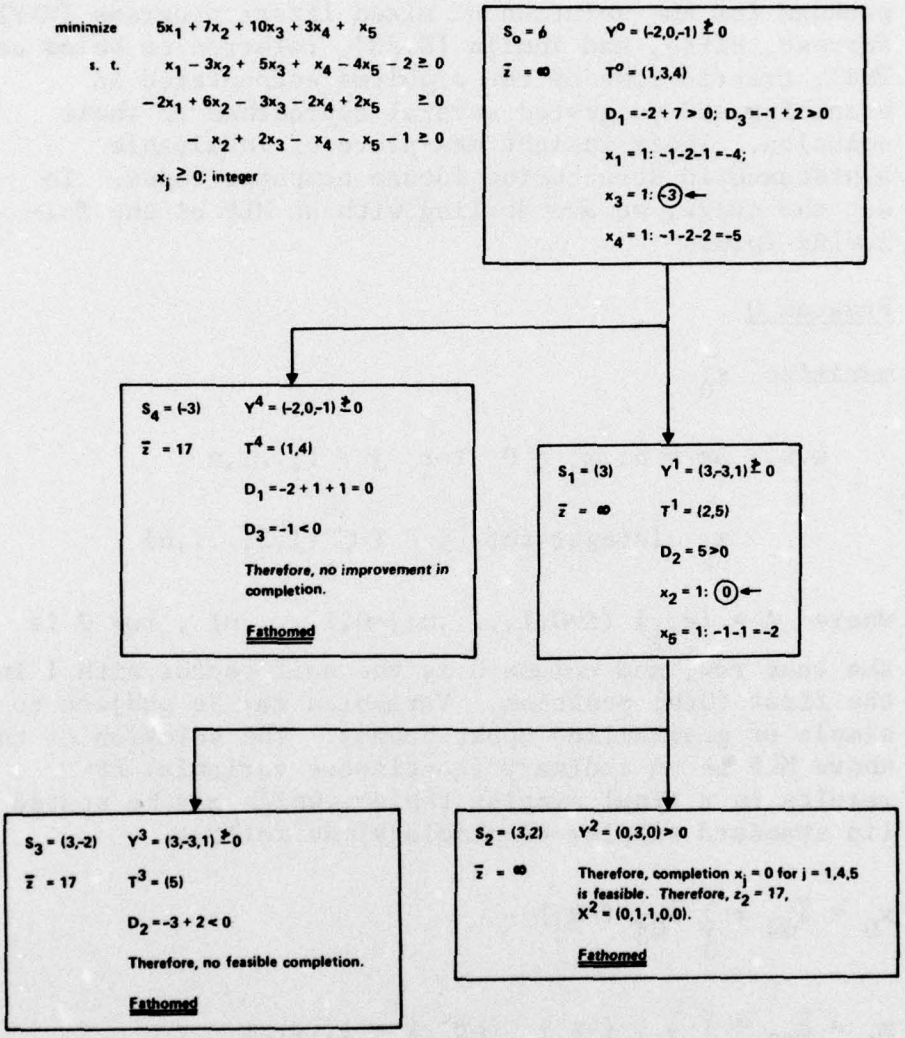


Figure 8.6 - Solution of an ILP.

their experience with UMPIRE, a proprietary computer package for the solution of mixed linear programs (MLP), Forrest, Hirsh, and Tomlin [8.23], referred to below as FH&T, treated some of the problems encountered in branching and suggested several approaches to their solution. Their insight may prove of invaluable assistance in structuring future computer codes. To set the stage, we are dealing with an MLP of the following form.

Program Q

maximize x_0

s.t. $Ax = b$, $x_j \geq 0$ for $j = 1,\ldots,n$

x_j integer for $j \in I \subset \{1,2,\ldots,n\}$

where $A = [a_{ij}]$ $(i=0,1,\ldots,m; j=0,1,\ldots,n)$, row 0 is the cost row, and column 0 is the unit vector with 1 in the first (0th) position. Variables may be subject to simple or generalized upper bounds. The solution of the above MLP as an ordinary (continuous variable) LP results in a final simplex tableau which may be stated (in standard simplex terminology) as follows.

$$x_0 = \bar{a}_{00} + \sum_j \bar{a}_{0j} (-x_j)$$

$$x_i = \bar{a}_{i0} + \sum_j \bar{a}_{ij} (-x_j) \quad \text{for} \quad i = 1,\ldots,m$$

By the simplex criterion for optimality, we have $\bar{a}_{i0} \geq 0$ and $\bar{a}_{0j} \geq 0$ for all i and j. For every $i \in I$, let

$$\bar{a}_{i0} = \left[\bar{a}_{i0}\right] + f_{i0},\ 0 \leq f_{i0} < 1$$

where $\left[\bar{a}_{i0}\right]$ is the largest integer not exceeding $\bar{a}_{i0}$.

The solution of Program Q by B&B involves maintaining a list of LP problems or subproblems derived from the original MLP, obtained by imposing tighter bounds on the integer variables and always recording the best integer solution obtained so far and its value x_0^c. The basic steps in branching are the following.

1. Problem (Node) Selection. Select a problem with fractional values from the list whose objective function satisfies $x_0 > x_0^c$. If none exists, terminate.

2. Choose A Branching Variable (the "Arbitrated" variable). Among the variables in the selected problem, choose a fractional variable for branching. Denote the chosen variable by x_p.

3. Branch (Arbitrate). Create two new subproblems (nodes) by adding the following new restrictions.

$x_p \leq \left[\bar{a}_{p0}\right]$ to yield subproblem p1

$x_p \geq \left[\bar{a}_{p0}\right] + 1$ to yield subproblem p2

The procedure is "straightforward" except for the fact that each step is in need of operational definition (which is the subject of the FH&T paper).

Consider Step 3 first. One may solve each descendant subproblem, or one may solve only one of the two LPs, postponing the solution of the other to later in the hope that it never need be done because of dominance considerations. Alternatively, one may "jockey" between the two new descendant subproblems p1 and p2; the authors refer to the alternation between the two nodes as "node swapping." For example, one may investigate the other subproblem as soon as the degradation in the first branch is found to exceed the "penalty" of the second as

explained below.

Three questions are to be resolved before Step 2 is successfully executed. How does one choose the branching variable? How does one decide on which branch to postpone (if branch postponement is desirable)? How does one bound its solution? To this end, FH&T discuss the following approaches.

(i) The Penalties Method. Penalties give a lower bound on the change (degradation) in the objective function as a consequence of forcing a currently non-integer variable to its adjacent integer values. From the theory of parametric LP it is evident that, assuming no change in basis, the imposition of a new ℓ.b. of $\left[\bar{a}_{p0}\right] + 1$ on x_p must decrease the objective function at least by the "up penalty"

$$U_p = \min_{j,\bar{a}_{pj}<0} (1-f_{p0})\,\bar{a}_{0j}/\left(-\bar{a}_{pj}\right)$$

If $\bar{a}_{pj} \geq 0$ for all j, let $U_p = \infty$ and x_p is a monotone decreasing variable. Similarly, the minimum "down penalty" incurred by placing an upper bound $\left[\bar{a}_{p0}\right]$ on x_p is given by

$$D_p = \min_{j,\bar{a}_{pj}>0} f_{p0}\,\bar{a}_{0j}/\bar{a}_{pj}$$

If $\bar{a}_{pj} \leq 0$ for all j, let $D_p = \infty$, and x_p is a monotone increasing variable. The penalties are lower bounds on the decrease in x_0 because we assumed no change in basis; hence the value of the objective function for these two branches must be bounded from above by $\bar{a}_{00} - U_p$, $\bar{a}_{00} - D_p$, respectively. If the descendant node is not dominated as a consequence of the new bound, then the above penalty calculations may be used

in the selection of the subproblem to be solved, postponing the other to later. Presumably, the subproblem giving the smallest degradation is the one selected.

The authors argue against this method since in many instances it leads to the wrong decision, thus prolonging the search. Garfinkel and Nemhauser [8.25] recommend the capitalization on monotone variables to reduce the size of the search tree. In particular, if row p is chosen as the partitioning row and x_p is monotone (increasing or decreasing), then it will have only one successor, and one need only to consider $x_p > \left[\bar{a}_p\right] + 1$ or $x_p > \left[\bar{a}_{p0}\right]$. They also point out, correctly, that the penalties U_p and D_p were derived without taking into account the integrality requirement on the nonbasic variables. Taking such requirements into account would generate new bounds on the penalties. For instance, in order to have an integer solution, some nonbasic variable must become positive and therefore not less than one. This immediately yields the ℓ.b. on penalty $\min_j a_{0j}$; which, incidentally does not depend on the partitioning row. One may wish to carry the idea of penalties a little further and determine a stronger bound on the penalty incurred. Two approaches suggest themselves.

(a) Assume that the current basis does not change, and determine a feasible solution when x_p is rounded up or down.

(b) Assume that the basis will change with the introduction of some nonbasic variable at a positive level. Determine the cheapest such transformation that retains primal feasibility, and its associated cost.

Naturally, the price paid for improved bounds is the additional computing. The efficacy of such approach is currently under investigation by the authors. Also see Breu and Burdet [8.6].

(ii) The Method of Priorities. Priorities are accorded to variables a priori, and are based primarily on the analyst's knowledge of the physical problem. Then one would select and branch on that variable not within some tolerance of an integer value (the "arbitration level") which has the highest priority.

(iii) The Best Projection Method. While this method is more appropriately related to Step 1 (Node Selection) rather than Step 2 (Variable Selection), it finds its place here because of the bearing it has on the next method of variable selection. The logical justification of this method is rather lengthy, albeit intuitively appealing. But its statement is rather simple. Suppose that an estimate of x_0^* can be made. Let s denote the "sum of integer infeasibilities,"

$$s \triangleq \sum_{j \in I} \min(f_{j0}; 1-f_{j0})$$

and let

$$\lambda \triangleq \frac{x_0^0 - x_0^*}{s^0} \geq 0$$

where x_0^0 is the value of the objective function of Program Q when solved as LP, and where s^0 is its corresponding sum of integer infeasibilities. (Note that the x_0's measure degradation from Program Q.) Then, for any node k with objective value x_0^k and sum value s^k, we define the "projection"

$$p^k = x_0^k - \lambda s^k$$

The best projection (BP) criterion for Step 1 of the B&B algorithm is now to choose the outstanding node with the largest value of p_k. The rationale for this is that p_k measures the approximate value of the integer solution we can expect to attain from node k.

The term "projection" stems from the fact that λ essentially projects the sum of integer infeasibilities s^k on the $s \mp 0$ axis in a particular direction (namely, a direction parallel to the line joining the points $\left(s_0^0, x_0^0\right)$ and $\left(0, x_0^*\right)$ in the $s - x_0$ domain).

(iv) The Pseudo-Costs Method. A close scrutiny of the definition of λ in the above method reveals that it can be interpreted as the "cost" of removing one unit of infeasibility. In fact, the last equation may be rewritten as

$$x_0^k - \lambda s^k = x_0^k - \sum_{i \varepsilon I} \min\{\lambda f_{i0}; \lambda(1-f_{i0})\}$$

Hence, in using this expression to estimate an obtainable integer solution we are implicitly "costing" the change in the variable at *the same cost* per unit change (whether up or down). Since this may not be true in general, we are led to the new estimate

$$e^k = x_0^k - \sum_{i \varepsilon I} \min\{d_i f_{i0}; u_i(1-f_{i0})\}$$

where d_i and u_i are the estimated costs per unit decrease or increase in variable x_i, respectively. The determination of the values d_i and u_i, as well as their revision as the iterations proceed, are discussed at length in the paper of FH&T.

(v) The Percentage-Error (P.E.) Method. We have the definition

$$\text{P.E.} = 100 \left(x_0^c - e^k\right) / \left(x_0^k - x_0^c\right)$$

for each node k. Essentially, it measures the degree of error if the current solution x_0^c is not optimal. If the P.E. is large and positive, it implies that a better integer solution is very unlikely to be found from this branch.

8.4 Bounding

We are always interested in the greatest lower bound (least upper bound) in the case of minimization (maximization). Unfortunately, this is oftentimes achieved at the heavy price of extensive calculation. Hence, there is the ever-present trade-off between a tight bound obtained at a considerable cost, and a loose one that is easily calculated. The only dictum that can be stated relative to this choice is that it may be worthwhile to put the effort in bounding nodes "higher up the tree," because then if fathomed we would save the enumeration of all their descendants. On the other hand, it is always advisable to obtain both upper and lower bounds on the value of z^* and preferably the tightest such bounds. Since a feasible solution always provides an u.b. (a ℓ.b.) in minimization (maximization) problems, it behooves the analyst to start the search procedure after having obtained as "good" a solution as possible, without the expenditure of an inordinate amount of effort in obtaining such a solution. Bounds on the value of the optimum are obtained by relaxing one constraint, or several constraints, of the original problem since the optimum of the relaxed problem is a bound on z^*. In certain instances the constraint to be relaxed is almost self-evident--such as relaxing the integer requirements in an ILP problem and solving as an ordinary LP. This is the relaxation adopted by Land and Doig [8.35] in their pioneering work. Oftentimes, though, the constraint to be relaxed requires insight into the problem to gain the "most mileage," by removing the more complicating constraint, without too much sacrifice in the value of the objective function.

We illustrate the concept of bounding with the following three examples. The first is, more or less, straightforward; the second exemplifies how a bound may be improved, that is, made tighter; the third exemplifies the need for the judicious choice of the constraint to be relaxed.

Example 8.4. Consider the problem of minimizing the total "makespan" in scheduling N jobs on three machines in series. We shall develop the ℓ.b. established by Lomnicki [8.38], which was apparently arrived

at independently by Ignall and Schrage [8.33]. It is well-known that for three machines an optimal schedule permits no "passing" and hence the order of jobs will be preserved on all three machines. Let $W = (w_1, w_2, \ldots, w_N)$ denote a permutation of the numbers $1,2,\ldots,N$ and let $f(w_i,j)$ represent the "earliest finish time" of job w_i on machine j, $j = 1,2,3$. Let the processing time of job w_i on machine j be denoted by $y(w_i,j)$. Then, clearly, for a given permutation W,

$$f(w_i,j) = \max\ [f(w_{i-1},j); f(w_i,j-1)] + y(w_i,j)$$

since to complete job w_i on machine j the time $y(w_i,j)$ must elapse after the machine became free from job w_{i-1}, or the job w_i became available from the previous machine $j-1$ whichever happens last. For ease of notation, denote $y(w_i,1)$ by $a(w_i)$, $y(w_i,2)$ by $b(w_i)$, and $y(w_i,3)$ by $c(w_i)$. Consider any partial schedule $w_1,w_2,\ldots,w_k$ which specifies the sequence of the first k elements, $k < N$. Then it is evident that the completion of all the remaining jobs consumes no less time than any of the following three values.

$$g' = f_W(w_k,3) + \sum_{j=1}^{N-k} c(w_{k+j})$$

$$g'' = f_W(w_k,2) + \sum_{j=1}^{N-k} b(w_{k+j}) + \min_{1\le j\le N-k} c(w_{k+j})$$

$$g''' = f_W(w_k,1) + \sum_{j=1}^{N-k} a(w_{k+j}) + \min_{1\le j\le N-k} [b(w_{k+j}) + c(w_{k+j})]$$

Lomnicki puts the lower bound $\max(g', g'', g''')$ on the partial schedule.

Example 8.5. Consider the problem of scheduling N jobs on M identical machines. Each job j has a fixed processing time y_j and a penalty coefficient p_j. A penalty $p_j t$ is incurred if job j is completed at time t (in other words, all jobs have a due date equal to zero, and they accumulate penalty starting from that time). Eastman, Even, and Isaacs [8.10], referred to below as EE&I, derived a lower bound on the optimum as follows. Let C_i be the symbol for the optimal cost of scheduling the N jobs on i machines, $1 \leq i \leq M$. Thus C_1 is the known optimal cost on one machine and C_N is the known optimal cost on N machines. In particular, the minimal cost (of scheduling N jobs on a single machine when a linear penalty is accumulated starting at time zero) is given by scheduling the jobs in their natural order, that is, in order of nonincreasing values of the ratio p_j/y_j. If the jobs are so numbered (if we have $p_1/y_1 \geq p_2/y_2 \geq \cdots \geq p_N/y_N$) then the minimal cost is given by

$$C_1 = \sum_j p_j \sum_{i \leq j} y_i$$

On the other hand, C_N is evidently given by $\sum_j p_j y_j$. Then EE&I assert that the desired lower bound on C_M is given by

$$\frac{1}{M}\left[C_1 + \frac{M-1}{2} C_N\right]$$

A sharper ℓ.b. was developed by Elmaghraby and Park [8.16] for the slightly more generalized problem in which each job j has a due date $d_j = y_j$. Their development was based on the remark that EE&I's ℓ.b. is based on the assumption that all the machines are

available at time zero. Clearly, given any partial schedule S_k of k jobs on the M machines, the times of earliest availability of each machine may be different from zero. This leads immediately to interest in developing a sharper ℓ.b. when the machines are available at time $T \geq 0$. To this end, let $C_{n,v}(T)$ denote the optimal cost of scheduling n jobs on v machines when all machines are available at time $T \geq 0$. Let T_j denote the time of completion of job j under the partial schedule S_k; m_ℓ is the last job on machine m; T_{m_ℓ} is the time of completion of the last job on machine m; $T_{min} = \min_m T_{m_\ell}$; and, $\bar{S}_k = N - S_k$. Then a ℓ.b. on the cost of the completion of S_k is given by

$$\sum_{j\in S_k} p_j(T_j - d_j) + \max\left\{0; \frac{1}{M} C_{k-N,1}(T_{min}) + \frac{M-1}{M} T_{min} \sum_j p_j - \frac{M-1}{M} C_{N-k,N-k}(0)\right\}$$

where $C_{N-k,N-k}(0) = \sum_{j\in \bar{S}_k} p_j y_j$. The first term is the cost incurred in scheduling the k jobs in S_k. The partial schedule S_k leaves the M machines with earliest availabilities $T_{1_\ell}, T_{2_\ell}, \ldots, T_{M_\ell}$. The smallest value of (earliest) availabilities is T_{min}, which was conservatively taken to be the availability of all machines. The second term is essentially EE&I's ℓ.b. corrected for the jobs already included in S_k and the earliest availability of the machines.

Example 8.6. In treating the problem of scheduling lots on a single facility over a finite horizon, Elmaghraby

and Mallik [8.14] addressed themselves to the following specific version. There are N items to be produced on the same facility. In any period (say, day or week), the facility may be devoted to the production of only one item. Item i is produced at the rate p_i per period, but is continuously consumed at the rate r_i per period where $r_i < p_i$. Given the initial "on hand" stock of each item, and the desired terminal inventory at the end of a planning horizon of length H, determine the minimum cost schedule (if one exists), where cost is defined in terms of inventory cost and back-order penalty, which vary from item to item. The constraint that Elmaghraby and Mallik chose to relax in their B&B approach is the noninterference constraint. Then the items are independent, which implies that the facility is devoted to the production of one item only. The determination of the optimal schedule under such an assumption is an intriguing problem in its own right, and was treated by Elmaghraby and Dix [8.15]. Fortunately, it proved to be of extremely simple form which requires a nominal amount of computing.

In relation to bounding, we wish to advance several dicta.

IV. The Use of the Previous ℓ.b. Calculations (for the Parent Node) as Lead-Off to the New ℓ.b. (of the Descendant Node). The pertinence of this concept increases with the amount of effort required in the calculation of the ℓ.b. The concept was used by Land and Doig [8.35] in their treatment of ILP, and by Elmaghraby [8.13] in his treatment of the problem of scheduling activities subject to resource constraints. In both instances the ℓ.b. at a node is determined by solving a (continuous) linear program. The ℓ.b. of a node is constructed from the parent optimal basis.

Another excellent example of capitalizing on the optimal solution of the previous iteration was provided by Srinivasan and Thompson [8.44] in their treatment of the traveling salesman problem (TSP) by the so-called "operator theory." The reader will recall that the TSP is a restricted assignment problem, restricted to assignments that are tours. Consequently, the search for the optimal tour in the TSP may be viewed as the

search for the optimal assignment (in the assignment problem) that is a tour. (This is not the only way to interpret the TSP. For example, Held and Karp [8.31] interpret it as a one-tree problem, hence the search for the optimal tour in the TSP is reduced to the search for the optimal one-tree. B&B methods are then used, and they report excellent computing results.) By utilizing the established properties of parametric linear programming, specialized to the assignment problem, the authors achieve the capability of probing the consequences of increasing a nonbasic variable (at the expense of a basic variable) in the optimal solution of the assignment problem, without in fact undertaking such changes. Because of the ease with which bounds can be established on such probes, the authors report excellent computing results.

V. Choose an Easy-To-Calculate Bound. The value of this concept rests on the fact that bounds are evaluated a large number of times over the life of a search, and if it is time-consuming it will render the search impractical.

Example 8.7. Nowhere is this concept more apparent than in the solution of the (linear) knapsack problem by the so-called "Greedy Algorithm." The setting of the problem is as follows.

$$\text{maximize} \quad \sum_i v_i x_i$$

$$\text{s.t.} \quad \sum_i a_i x_i \leq b \ , \ x_i = 0,1 \quad \text{for all} \quad i$$

and where the v_i , the a_i , and b are given positive integer constants. The rationale for the Greedy Algorithm is the well-known observation that, in the absence of the integer requirements, the optimum is readily obtained by renumbering the variables in order of non-increasing v_i/a_i , and putting $x_j = \min \left(1, b - \sum_{i=1}^{j-1} a_i\right)$, $j = 1,2,\ldots,n$. The Greedy Algorithm

approach to the knapsack problem proceeds in exactly the same fashion, but then it branches on the fractional variable, denoted by x_r, thus creating two subsets corresponding to $x_r = 0$ and $x_r = 1$. Of these two nodes, we investigate (branch on) the node with the best (fractional) solution, which is an u.b. on the optimal value. In case of ties, apply any heuristic to break it, such as random choice among the tied nodes. Continue the process of dichotomizing the solution space; each time an integer solution is achieved it provides a new ℓ.b. on the value of the optimum (if it is better than the incumbent). If all the u.b.'s of the unbranched-from nodes are smaller than the current (integer) ℓ.b., it is also optimal. Otherwise, branching and bounding continues from other nodes until the optimum is achieved.

To illustrate, consider the following knapsack problem.

$$\text{maximize} \quad z = 5x_1 + 4x_2 + 3x_3 + 2x_4$$

$$\text{s.t.} \quad x_1 + 2x_2 + 3x_3 + 4x_4 \leq 5$$

$$x_j = 0,1 \quad \text{for all} \quad j .$$

The search tree is shown in Figure 8.7. In this example we explicitly enumerated only 8 solutions out of the possible 16 solutions. In larger problems, the savings are, fortunately, more pronounced than in this small example.

VI. Relax the Objective Function Instead of a Constraint. Sometimes bounds are easily computed by relaxing the form of the objective function, rather than a constraint. Perhaps the following example illustrates this concept best.

Example 8.8. In the field of project planning and control, a problem that has been extensively studied is that of reducing the duration of a project (the so-called project "crashing" or "compression") at minimal cost. The optimal reduction under the assumption of

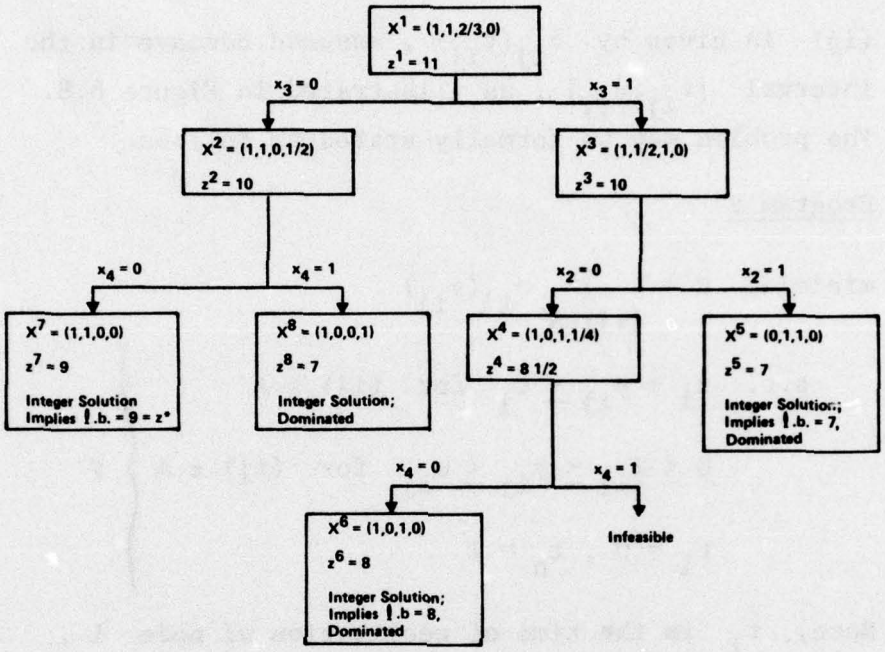


Figure 8.7 - Search tree for a knapsack problem.

linear or convex costs for the individual activities has been treated by several researchers, see Clark [8.7], Elmaghraby [8.12], and Fulkerson [8.24], among others, who used algorithmic approaches. However, the case of concave cost functions was treated by Falk and Horowitz [8.21] using a rather ingenious B&B approach, first proposed by Rech and Barton [8.43]. Their basic idea was to relax the objective function into the largest linear function that underestimates the concave cost of each activity. This reduces the problem to an LP problem, which can be easily solved; whose optimum is a ℓ.b. on the optimum cost desired. Also, by virtue of it being a feasible solution to the original problem, it also provides an u.b. More importantly, the LP solution suggests the partitioning of the solution space, the branching process, which is proved to terminate in a finite number of steps. Briefly, the procedure is as follows.

Let the duration of activity (ij) be denoted by y_{ij}, and assume its upper and lower limits are denoted by u_{ij} and ℓ_{ij}, respectively. The cost of activity

(ij) is given by $c_{ij}(y_{ij})$, assumed concave in the interval $[\ell_{ij}, u_{ij}]$, as illustrated in Figure 8.8. The problem may be formally stated as follows.

Program P

$$\text{minimize} \quad C = \sum_{(ij)\in A} c_{ij}(y_{ij})$$

$$\text{s.t.} \quad \left.\begin{array}{l} t_i + y_{ij} \le t_j \quad \text{for} \quad (ij) \in A \\ 0 \le \ell_{ij} \le y_{ij} \le u_{ij} \quad \text{for} \quad (ij) \in A \\ t_1 = 0 \ , \ t_n = T \end{array}\right\} F$$

Here, t_i is the time of realization of node i , A is the set of activities (arcs), and T is the specified duration of the projects. The approach is simply the following. Suppose that the concave cost function of Figure 8.8(a) is approximated by a linear function as shown by the dotted line in (b), which is the highest linear function which underestimates $c_{ij}(y_{ij})$. Denote such a linear cost function by $c^1_{ij}(y_{ij})$. Then we may take $C^1(Y)$ as a first (lower) approximation to the objective function of the Program P, and formulate the first "estimating problem" Q^1 as follows.

Program Q^1

$$\text{minimize} \quad C^1(Y) = \sum_{(ij)\in A} c^1_{ij}(y_{ij})$$

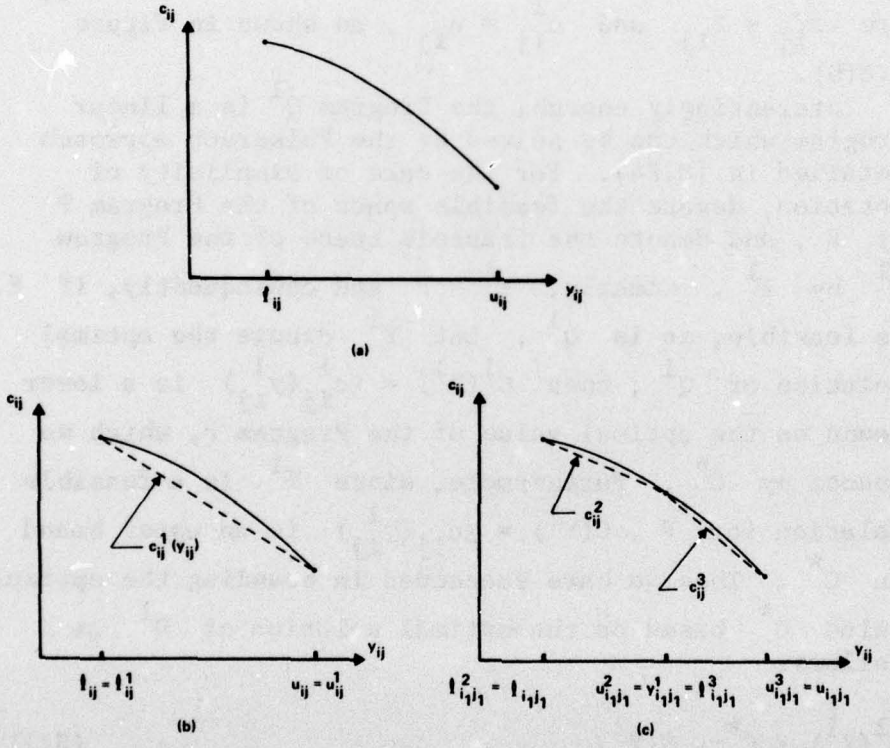


Figure 8.8 - Approximating a concave cost function.

$$\left.\begin{array}{ll} \text{s.t.} & t_i + y_{ij} \leq t_j \quad \text{for} \quad (ij) \in A \\ & \ell_{ij}^1 \leq y_{ij} \leq u_{ij}^1 \quad \text{for} \quad (ij) \in A \\ & t_1 = 0 \ , \ t_n = T \end{array}\right\} F^1$$

Here, the upper and lower bounds on the duration y_{ij} are $\ell_{ij}^1 = \ell_{ij}$ and $u_{ij}^1 = u_{ij}$, as shown in Figure 8.8(b).

Interestingly enough, the Program Q^1 is a linear program which can be solved by the Fulkerson approach detailed in [8.24]. For the sake of simplicity of notation, denote the feasible space of the Program P by F, and denote the feasible space of the Program Q^1 by F^1. Clearly, $F^1 = F$ and consequently, if P is feasible, so is Q^1. Let Y^1 denote the optimal solution of Q^1; then $C^1(Y^1) = \Sigma c_{ij}^1(y_{ij}^1)$ is a lower bound on the optimal value of the Program P, which we denote by C^*. Furthermore, since Y^1 is a feasible solution for P, $C(Y^1) = \Sigma c_{ij}(y_{ij}^1)$ is an upper bound on C^*. Thus we have succeeded in bounding the optimal value C^* based on the optimal solution of Q^1 as follows.

$$C^1(Y^1) \leq C^* \leq C(Y^1) \tag{8.1}$$

Clearly, if equality holds throughout (8.1) then the current trial solution is optimal. This remark holds for all subsequent iterations, and hence will not be repeated. Now suppose that strict inequality holds above; then there is room for improvement. This is accomplished by producing a closer (albeit still an underestimating) linear approximation to the cost function $c(y)$. Consider the difference

$$c_{ij}(y_{ij}^1) - c_{ij}^1(y_{ij}^1) \tag{8.2}$$

which is always ≥ 0 and suppose that activity $(i_1 j_1)$ yields the maximum such difference, that is,

$$c_{i_1 j_1}\left(y^1_{i_1 j_1}\right) - c^1_{i_1 j_1}\left(y^1_{i_1 j_1}\right) = \max_{(ij)\in A} \left[c_{ij}\left(y^1_{ij}\right) - c^1_{ij}\left(y^1_{ij}\right)\right] > 0$$

Such an activity must exist, for otherwise the differences of (8.2) are zero for all activities, implying equality in (8.1) which is a contradiction. (In case of ties, any tied activity will do.) Divide the feasible domain of $y_{i_1 j_1}$ into two subintervals: $\left[\ell_{i_1 j_1}, y^1_{i_1 j_1}\right]$ and $\left[y^1_{i_1 j_1}, u_{i_1 j_1}\right]$. (Recall that $y^1_{i_1 j_1}$ is the value of the duration of activity $(i_1 j_1)$ obtained from the optimal solution of the Program Q^1.) Construct the two cost functions: $c^2_{i_1 j_1}\left(y_{i_1 j_1}\right)$ and $c^3_{i_1 j_1}\left(y_{i_1 j_1}\right)$ as the maximum linear approximations which underestimate the original cost function $c_{i_1 j_1}\left(y_{i_1 j_1}\right)$ in the two subintervals $\left[\ell_{i_1 j_1}, y^1_{i_1 j_1}\right]$ and $\left[y^1_{i_1 j_1}, u_{i_1 j_1}\right]$ illustrated as in Figure 8.8(c). Now define two linear programs as follows.

Program Q^2

$$\text{minimize} \quad \sum_{(ij)\neq(i_1 j_1)} c^1_{ij}(y_{ij}) + c^2_{i_1 j_1}\left(y_{i_1 j_1}\right)$$

$$\left.\begin{array}{ll} \text{s.t.} & t_i + y_{ij} \leq t_j \quad \text{for all} \quad (ij) \in A \\ & \ell_{ij} \leq y_{ij} \leq u_{ij} \quad \text{for} \quad (ij) \neq (i_1 j_1) \\ & \ell_{i_1 j_1} = \ell^2_{i_1 j_1} \leq y_{i_1 j_1} \leq u^2_{i_1 j_1} = y^1_{i_1 j_1} \\ & t_1 = 0 \ , \ t_n = T \end{array}\right\} F^2$$

Program Q^3

$$\text{minimize} \quad \sum_{(ij)\neq(i_1 j_1)} c^1_{ij}(y_{ij}) + c^3_{i_1 j_1}\left(y_{i_1 j_1}\right)$$

$$\left.\begin{array}{ll} \text{s.t.} & t_1 + y_{ij} \leq t_j \quad \text{for all} \quad (ij) \in A \\ & \ell_{ij} \leq y_{ij} \leq u_{ij} \quad \text{for} \quad (ij) \neq (i_1 j_1) \\ & y^1_{i_1 j_1} = \ell^3_{i_1 j_1} \leq y_{i_1 j_1} \leq u^3_{i_1 j_1} = u_{i_1 j_1} \\ & t_1 = 0 \ , \ t_n = T \end{array}\right\} F^3$$

The logic of these two programs rests on the observation that the duration of the activity $(i_1 j_1)$ must lie in either of the two subintervals of Figure 8.8(c). Note that the only difference between these two programs is in the definition of the range of the duration $y_{i_1 j_1}$.

Both programs are feasible since the point Y^1 is still a feasible point of either of them. Let F^2 denote the space of feasible solutions for the Program Q^2 and let F^3 denote the space of feasible solutions for the Program Q^3. Clearly, F^2 and F^3 are defined by

$$F^2 = F^1 \cap \left\{(Y,t):\ \ell^1_{i_1j_1} \le y_{i_1j_1} \le y^1_{i_1j_1}\right\}$$

$$F^3 = F^1 \cap \left\{(Y,t):\ y^1_{i_1j_1} \le y_{i_1j_1} \le u_{i_1j_1}\right\} \tag{8.3}$$

Furthermore,

$$c^2_{ij}(y_{ij}) = c^3_{ij}(y_{ij}) = c^1_{ij}(y_{ij}) \quad \text{for} \quad (ij) \neq (i_1j_1) \tag{8.4}$$

Therefore, Problems Q^2 and Q^3 may be succinctly stated as:

Program Q^2: minimize $C^2(Y)$ s.t. $(Y,t) \in F^2$

Program Q^3: minimize $C^3(Y)$ s.t. $(Y,t) \in F^3$

Now both problems Q^2 and Q^3 may be solved by the Fulkerson algorithm, yielding optimal durations Y^2 and Y^3, respectively. By virtue of the fact that the cost functions C^2 and C^3 serve as tighter underestimates of C over their domains, and that the feasible space $F \equiv F^1 = F^2 \cup F^3$, we have

$$C^1(Y^1) \le \min\left\{C^2(Y^2), C^3(Y^3)\right\} \le C^* \le \min\left\{C(Y^1), C(Y^2), C(Y^3)\right\} \le C(Y^1)$$

The rightmost inequality follows from (Y^2,t^2) and (Y^3,t^3) being feasible solutions to Problem P. We have thus achieved improved bounds on the optimal value C^*.

From this point onwards, the algorithm proceeds in a series of stages. The zeroth stage, just detailed above, consists of Problem Q^1 and its solution Y^1. The first stage consists of Problems Q^2 and Q^3 and their solutions Y^2 and Y^3. The kth stage consists of Problems Q^{2k} and Q^{2k+1} and their solutions. The process is conveniently depicted by a typical binary-search tree whose nodes correspond to the Problems Q^S. Figure 8.9 depicts such a tree with four stages and nine nodes. Branching occurs when a particular activity is selected to have one of its (duration) intervals divided into two subintervals, as exemplified in Figure 8.8(c) with costs and bounds redefined as in (8.3) and (8.4). A branching node s may be selected according to some heuristic rule; for example, it may be done by choosing that Problem Q^S whose optimal value $C^S(Y^S)$ is minimal over all cost values associated

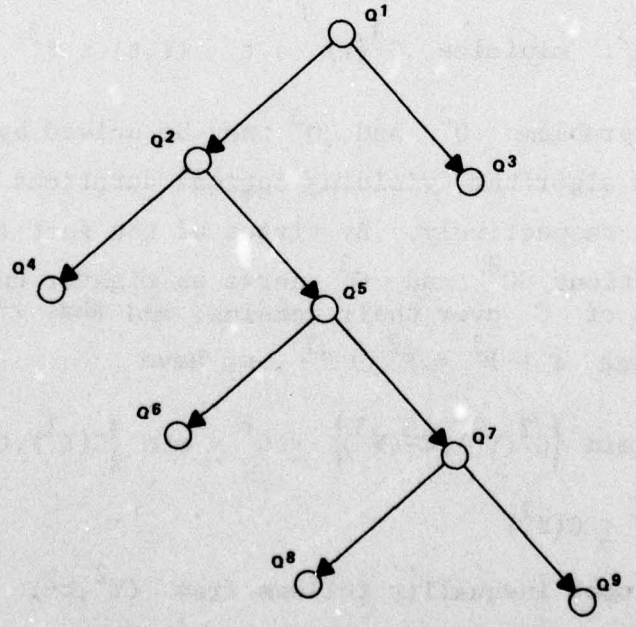


Figure 8.9 - A binary search tree.

with intermediate nodes (nodes from which no branching has occurred, denoted by $I(k)$ at stage k). The rationale here is that Problem Q^s has the smallest lower bound on C^* and, hopefully, the feasible space F^s will contain a point yielding a value to C approximately equal to $C^s(Y^s)$. Another possible heuristic rule is as follows: choose that Problem Q^s whose optimal solution Y^s yields the smallest interval of uncertainty on C^*. In any event, having chosen a Problem Q^s at stage k from which to branch, we create the Problems Q^{2k} and Q^{2k+1} in a manner exactly analogous to the manner in which Q^2 and Q^3 were created from Problem Q^1. The feasible spaces F^{2k} and F^{2k+1} are thus defined by

$$F^{2k} = F^s \cap \left\{(Y,t): \ell_{i_s j_s} \leq y_{i_s j_s} \leq y^s_{i_s j_s}\right\}$$

$$F = F^s \cap \left\{(Y,t): y^s_{i_s j_s} \leq y_{i_s j_s} \leq u_{i_s j_s}\right\}$$

Furthermore, $c^{2k}_{i_s j_s}$ is the highest linear function underestimating $c_{i_s j_s}$ over the subinterval $\left[\ell_{i_s j_s}, y^s_{i_s j_s}\right]$ while $c^{2k+1}_{i_s j_s}$ is the highest linear function underestimating $c_{i_s j_s}$ over subinterval $\left[y^s_{i_s j_s}, u_{i_s j_s}\right]$. The sets F^{2k} and F^{2k+1} are feasible, since the point (Y^s, t^s) lies in both sets. The new problems Q^{2k} and Q^{2k+1} are linear problems with network constraints similar to those of Problem P and thus

may be solved by the Fulkerson algorithm. The optimal value C^* is found at stage $k+1$ by

$$v^{k+1} = \min_{s \in I(k+1)} \{C^s(Y^s)\} \leq C^* \leq \min_{s=1,2,\ldots,2k+1} \{C(Y^s)\}$$

$$= w^{k+1}$$

The process continues until either $v^k = w^k$ at some stage k, or when the interval bounding C^* is deemed small enough for practical purposes.

Three remarks should now be made. First, the choice of the subintervals on the duration $y_{i_s j_s}$ of Problem Q^s is arbitrary; the division indicated above seems to be a reasonable heuristic. Second, the above approach is clearly applicable to piecewise linear functions, whether concave or convex; in fact, the numerical example worked out in Falk and Horowitz [8.21] contained arcs in both categories. Third, the proof that the algorithm is finite rests on the fact that the function C is concave and defined over a convex polytope: it must assume its minimum at one of the finite vertices of the feasible space F. Most of the subproblems Q^s have their solutions at vertices of F. But since new vertices are created when new upper and lower bounds u^s_{ij} and ℓ^s_{ij} are added, some Q-problems may have solutions at points which are not vertices of F. The proof that only a finite number of such problems exist is given in Reference [8.22].

VII. Local Optima Provide Excellent Bounds. The dictum seems pedantic, yet its application, where possible, yields significantly improved results. To some degree this dictum, which advises the analyst to seek the optimum of the subspace under consideration, seems to be antithetical to Dictum V which advises against such optimization. Nevertheless, this dictum should be taken to read: if one can easily derive a local

optimum, then it is worth implementing.

Example 8.9. One of the more recent illustrations of this dictum was provided by Mitten and Tsou [8.41] referred to below as M&T. They addressed themselves once more to the problem of scheduling N tasks on a single facility to minimize an objective cost expression. Their approach was to combine simple bounding with local optimality to achieve efficiency and fast convergence. We introduce their terminology and notation.

S : a finite set of N elements (the tasks)

$\sigma \subset s$: a subset of elements of S (a subset of the tasks)

P_σ : the set of all permutations of σ, with $P = \bigcup_\sigma P_\sigma$

$p \in P$: a permutation in P, $p = (w_1, w_2, \ldots, w_n)$, $n \leq N$, in which w_i is the task occupying the ith position in the sequence

S_p : the unordered set of elements contained in p, and $\bar{S}_p = S - S_p$

p^m : $(w_1, w_2, \ldots, w_m)$, the set of the first m elements of $p \in P$ with p^0 the null permutation

(p,q) : a permutation formed by the concatenation of the two dispoint permutations p and q, both in P

p_i : the ith element in the permutation $(w_1, w_2, \ldots, w_m)$

To each element $x \in S$ there are two given real and finite constants: $c_x > 0$, denoting the cost of task

AD-A050 798 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 15/5
MODERN TRENDS IN LOGISTICS RESEARCH. PROCEEDINGS OF A CONFERENC--ETC(U)
1976 W H MARLOW N00014-75-C-0729
UNCLASSIFIED NL

3 OF 5
AD
A050798



x discounted to its start time, and $d_x \geq 0$ denoting its duration. For any permutation $p \in P$, let

$$D_p = \sum_{x \in S_p} d_x \quad \text{with} \quad D_{p_0} = 0$$

Assume that "now" is the start time of the scheduling process. Let $f(\xi) \geq 0$ be a real-valued bounded function, say a discount factor. Let $C(p,t)$ denote the cost associated with the permutation p when initiated at time t. For any real constant D_0, $0 \leq D_0 < \infty$, the cost associated with permutation $p \in P$ is given by

$$C(p,D_0) = \sum_{m=1}^{n} c_{p_m} f\left(D_0 + D_{p_{m-1}}\right)$$

Note that

$$C[(p,q),0] = C(p,0) + C(q,D_p)$$

The optimum is defined by

$$C^*(D_0) = \min_{p \in P_S} C(p,D_0)$$

which corresponds to some $p^* \in P^*(D_0)$.

Let us now turn to the generation of bounds. It is well known that a condition for local optimality of a permutation is that any contiguous binary switch (CBS) does not lead to improved objective value. The statement of this condition in M&T terminology is as follows. Let q and r be two disjoint permutations not containing the two distinct elements x and y. Let $p = (q,x,y,r)$ and $p' = (q,y,x,r)$; that is, p' is the permutation p with the elements x and y interchanged. Then

$$C(p,0) - C(p',0) = c_y[f(D_q+d_x) - f(D_q)] - c_y[f(D_q+d_y) - f(D_q)]$$

Let $R_w(D) = [f(D+d_w) - f(D)]/c_w$ for any $w \in S$ and $0 \leq D < \infty$. Then it can be seen that

$$R_x(D_q) \leq R_y(D_q)$$

is equivalent to

$$C(p,0) \leq C(p',0)$$

from which M&T obtain the following CBS condition

$$p \in P^* \text{ only if } R_x(D_q) \leq R_y(D_q) \tag{8.5}$$

Furthermore, p is locally optimum if condition (8.5) holds for every adjacent pair of elements in p. This condition leads immediately to the following construction to establish an upper bound B on the optimal value. Construct a complete permutation $g = (g_1, g_2, \ldots, g_N)$ by iteratively choosing g_i satisfying

$$R_{g_1}(0) \leq R_x(0) \quad \text{for all } x \in S$$

$$R_{g_i}\left(D_{g^{i-1}}\right) \leq R_x\left(D_{g^{i-1}}\right) \quad \text{for all } x \in S - g^{i-1}$$

Clearly, the value of any complete permutation is an upper bound; but this value $C(g,0)$ of the particular permutation g is usually a very good upper bound, if not optimal value. It is equally well-known that the cost of any partial permutation p is itself a lower bound b on the value of the optimum. However, if $f(\xi)$ is monotone in ξ, which is usually the case, we can do better. For, let σ be a nonempty proper subset of S which contains no elements of p.

Suppose that $f(\xi)$ is nonincreasing in ξ (the case of a discount factor). Consider the two partial permutations u and v defined as follows

$$u = (u_1, u_2, \ldots, u_n) \in P \quad \text{with} \quad c_{u_1} \leq c_{u_2} \leq \cdots \leq c_{u_n}$$

$$v = (v_1, v_2, \ldots, v_n) \in P \quad \text{with} \quad d_{v_1} \geq d_{v_2} \geq \cdots \geq d_{v_n}$$

Then a sharper ℓ.b. is given by

$$b(D_p) = C(p,0) + \sum_{i=1}^{n} c_{u_i} f\left(D_p + D_{v^{i-1}}\right)$$

In case $f(\xi)$ is monotone nondecreasing, reverse the order of elements in both u and v.

Thus the u.b. is established on the basis of local optima. The ℓ.b. is based on a well-known inequality that presumes the relaxation of the parameters of the problem. The reader has just been introduced to a third form of relaxation!

8.5 Dominance and Feasibility

The ultimate in efficiency of search techniques is to be able to rule out, or eliminate from the set of contenders, all points but one based on logical arguments alone. Unfortunately this happy state of affairs rarely occurs, and when it does, it rules out the very need for an iterative search procedure such as discussed here under B&B. Still, the concept is appealing and eminently useful. We have already witnessed such elimination (fathoming) through the use of bounds. The two other avenues are: the establishment of dominance relations between subsets of the feasible space, and the establishment that certain completions of partial solutions are infeasible. This latter consideration arises as a consequence of the fact that, more often than not, the original space being searched is "enriched" through the inclusion of infeasible points. Thus the words "dominance and infeasibility" refer to the dual activity of weeding out infeasible points and, among the feasible ones, demonstrating that some subset is

preferred to another.

The above use of dominance and infeasibility is essentially a process of elimination. Such elimination is motivated by one of two considerations: by inclusion, or by exclusion and decomposition and it is accomplished through the use of cuts, surrogate constraints, and relations. The following discussion elaborates on these notions.

VIII. Adaptive Cuts: Value Cuts and Configuration Cuts. By a "cut" is meant an additional constraint to the original statement of the problem. Such cuts are usually "adaptive" because their form and content depend on the stage of iteration and the particular partial solution being considered. The concept is based on the elementary observation that having arrived at a partial solution the analyst should, and usually can, augment the set of constraints on the basis of additional information available. These additional constraints "cut out" portions of the original feasible space. This results in sharper (lower or upper) bounds and reduced search effort. The concept is not new: it was first used by Little et al. [8.37] in their solution of the TSP. For instance, they asserted that given a partial-permutation $(1,i_2,i_3,\ldots,i_k)$ of the first k cities in a tour, then any completions which contain a city i_r in the subset of cities $1,i_2,\ldots,i_k$ are inadmissible; that is, such completions are "cut out" of the feasible space since the complete permutation would then contain a subtour. The generation of cuts in B&B procedures demands ingenuity and insight into the problem since the nature and form of cuts vary from problem to problem.

In general, there are two kinds of cuts: "value cuts" and "configuration cuts." The former type of cut is generated from knowledge of the value of the objective function, or knowledge of its bounds. The latter is gleaned from either the constraining set of equations or the set K of "other" restraints. Value cuts are illustrated by the following two examples.

Example 8.10. Consider the following general MLP problem which crops up regularly in the field of facility

location-and-allocation problems:

Program L

minimize $\phi = fy + cx$

s.t. $A_1 y + A_2 x = b$

$x \geq 0$, $y_i = 0,1$ and $y \in K$

Here, f , c are given vectors, A_1 and A_2 are given matrices, and K represents additional constraints on the y_i . For a physical interpretation of this model, one may consider the binary variable y_i to correspond to the dichotomy: have a facility in location i ($y_i=1$) or do not ($y_i=0$) . The fixed cost for establishing the facility is f_i independent of its size. The vector x may represent a level of the activities, with corresponding cost vector c . The matrices A_1 and A_2 are the coefficients of the variables in the constraining equations. The set K represents, for example, the so-called "bunching constraints" of the form

$$\sum_{i \in S_k} y_i = 1$$

which reflect the need for one, and only one, facility in a given subset of locations S_k . For ease of notation, denote by B the set of additional constraints imposed on the y and x vectors

$$B \triangleq \{y,x: x \geq 0 , y_i = 0,1 , y \in k\}$$

To illustrate the concept of value cuts we proceed to establish a sequence of lower and upper bounds on costs.

(a) A ℓ.b. on the total cost (LBTC). Solve the LP

$$\text{minimize} \quad \psi = fy + cx$$

$$\text{s.t.} \quad A_1 y + A_2 x = b$$

$$y \, , \; x \in B'$$

$$B' \triangleq \{y,x: \; x \geq 0 \; , \; 0 \leq y_i \leq 1 \; , \; y \in K\}$$

In other words, the LBTC is obtained by solving the original problem but with the integer requirements on the y_i relaxed.

(b) A ℓ.b. on the cost associated with the continuous part (LBCC). Solve the MLP

$$\text{minimize} \quad \gamma = cx$$

$$\text{s.t.} \quad A_1 y + A_2 x = b$$

$$x, \; y \in B$$

Note that here we are demanding that y_i be a binary variable. In some cases the retention of this requirement poses no computational difficulty (see [8.17]); otherwise, substitute the set B' for B.

(c) An u.b. on the cost associated with the integer part (UBIC)

If at any stage of the search tree there is available an estimate of the upper bound on the total cost (UBTC), such as a complete feasible solution, it can be utilized to calculate an UBIC as follows. Define TIC (TCC) as the total integer cost (total continuous cost) associated with any solution vector y of the integer part. Clearly, we are interested in considering only the vectors y which could improve the current UBTC, that is, those which yield

$TIC \leq UBTC - TCC$

Since $TCC \geq LBCC$, then, a fortiori,

$TIC \leq UBTC - LBCC$

and the quantity UBTC-LBCC is an u.b. on the value of the integer costs (UBIC). We have generated a value cut of the form

$fy \leq UBIC \ (=UBTC-LBCC)$

This cut is updated once a better UBTC and/or a better LBCC is obtained. The cut is incorporated, upon its generation, in the restraining set.

Example 8.11. The second form of value cuts is exemplified by the well-known Benders' cuts [8.5] which are based on a partitioning approach. Because of their frequent use in MLP problems, we briefly review their construction in the context of the above facility location MLP. Consider the mixed Program L stated at the beginning of Example 8.10 and let y^0 be any non-negative integer vector. The resulting problem is an ordinary (continuous) LP.

Program LC

$$\text{minimize} \quad \phi(y^0) = fy^0 + cx$$

$$\text{s.t.} \quad A_2x = b - A_1y^0$$

$$x \geq 0$$

The dual is

Program DLC

$$\text{maximize} \quad \psi(y^0) = fy^0 + u(b-A_1y)$$

$$\text{s.t.} \quad uA_2 \leq c$$

$$u \geq 0$$

If DLC has no feasible solution, then the vector y^0 is inadmissible. On the other hand, if DLC has an unbounded solution, then the primal LC is infeasible for all x, and again y^0 is declared inadmissible. This leaves y^0 admissible only if DLC possesses a finite maximum. Assume that to be the case, and suppose that it corresponds to $u^0(y^0)$. Clearly

$$\phi(y^0) = \psi^*(y^0) \geq \phi^*$$

and we may write

$$g^* = \min fy + \max u(b-A_2y)$$

where "min" is s.t. $y = 0,1$, and admissible, and "max" is s.t. $uA_2 \leq c$, $u \geq 0$. If T denotes the set of extreme points of the constraint set of (DLC) then T is of finite cardinality and

$$\psi^*(y^0) = fy^0 + \max_{u \in T} u(b-A_1y^0)$$

In order to restrict attention only to admissible y, we must guarantee that the solution to DLC is bounded. Suppose that it is not. Then it must be true that $u(b-A_1y^0)$ increases along some extreme ray. In other words, there would exist an extreme point u' and a direction v such that every point on the extreme ray

$$u' + \theta v, \quad \theta \geq 0$$

is feasible for DLC and $(u'+\theta v)(b-A_1y^0)$ increases with θ, or $\theta v(b-A_1y^0)$ increases with θ, implying that $v(b-A_1y^0) > 0$. To prevent this eventuality, let

$R = \{v: u + \theta v,\ \theta \geq 0$ is an extreme ray for some $u \in T\}$

Then we require y^0 to satisfy

$v(b-A_1y^0) \leq 0$ for every $v \in R$

Incorporating this into the original MLP we obtain

$g^* = \min fy + \max u(b-A_1y)$

where "min" is s.t. $y = 0,1$, and "max" is s.t. $u \in T$ and $v(b-A_1y) \leq 0$ for every $v \in R$. The original MLP may be transformed into an equivalent MLP by introducing the objective function of DLC as a variable ψ such that

$\psi \geq fy^0 + u(b-A_1y^0)$ for $u \in T$.

Then the equivalent MLP is given by

Program EL

$\phi^* = \text{minimum } \psi$

s.t. $\psi \geq fy + u(b-A_1y)$ for $u \in T$

$0 \geq v(b-A_1y)$ for $v \in R$

$y = 0,1$

This "mixed" program has all integer variables but one, namely ψ. It is usually approached as an ILP with the understanding that any algorithm for ILPs would have to be modified slightly to solve it. Essentially, the MLP of L has been partitioned into the continuous linear program LC and the "almost integer" linear program EL. Theoretically, EL can be solved only if all

the extreme points and extreme rays of DLC are known. Practically, EL can be solved iteratively by generating constraints only when they are needed.

Let $\bar{\phi}$ (possibly $+\infty$) be an upper bound on ϕ^*, obtained perhaps by a feasible solution. Let T^k and R^k be any subsets of extreme points and directions of extreme rays of $\{u: uA \leq c\ ,\ u \geq 0\}$ known at iteration k (at the start, both T^0 and R^0 are empty). Since $\phi^* \leq \bar{\phi}$ and

$$\phi^* \leq fy + u(b-A_1y) \quad \text{for every} \quad u \in T^k$$

then we may impose the constraint:

$$\bar{\phi} \geq fy + u(b-A_1y) \quad \text{for every} \quad u \in T^k$$

which replaces the first set of constraints in EL. To restrict the enumeration to admissible y, we further have

$$v(b-A_1y) \leq 0 \quad \text{for every} \quad v \in R^k$$

Equivalently, we have

$$(-f+uA_1)\ y \geq -\bar{\phi} + ub \quad \text{for every} \quad u \in T^k$$

$$vA_1y \geq vb \quad \text{for every} \quad v \in R^k$$

$$y = 0,1$$

The sets T^k and R^k increase in size (implying additional constraints) as iterations proceed, which is one possible drawback of the Benders' partitioning algorithm. Of course, the number of iterations is limited by 2^n, the number of possible binary values of the vector y. However, empirical evidence shows that termination is achieved long before the sets

T^k and R^k contain all feasible extreme points and extreme rays.

Configuration cuts are usually implicitly embedded in the definition of the problem, residing either in the constraints set or in the definition of the set K. However, if a cut in the solution space of the integer is sought, these implied restraints must be made explicit. For instance, in capacitated plant location problems, the set of constraints may specify that all demands at the various destinations must be met. Thus the following constraint is implied

$$yq \geq \text{RTDMN}$$

where q is the vector of capacities at the various locations, and RTDMN is the residual total demand. The efficiency of this cut is data-dependent, but it has proved to be most powerful in some applications [8.17]. For another instance, the set K implies mainly the "multiple choice" constraints of the form

$$\sum_{i \in S_k} y_i \leq 1 \quad \text{for all} \quad S_k \,, \; k = 1,2,\ldots$$

where S_k is a set of possible choices for a facility. These "bunching constraints" are adaptively updated as the search proceeds and partial solutions are obtained.

IX. Surrogate Constraints. Surrogate constraints do not augment the set of restraining equations, but rather substitute for them. The surrogate usually increases the search space. Hopefully, the solution of the problem with the surrogate is much easier than the original problem, which more than compensates for the loss in efficiency due to the expansion of the search space. Of course we are interested in the "tightest" surrogate, in the sense of minimal enlargement of the search space. If that is achieved, the resultant would be a tighter bound, which is always desirable.

Example 8.12. To fix ideas, consider the following ILP where $c \leq 0$ and $b \geq 0$.

Program R

maximize $z = cx$

s.t. $Ax \leq b$

$x = 0,1$

Suppose that at the kth stage of iteration we have a partial solution: some of the variables fixed at 0 form the set S_k^0 ; others fixed at 1 form the set S_k^1 ; and the remaining variables are still free and form the set F_k . At that particular node in the search tree, the problem may be stated as follows.

Program R_k

$$\text{maximize} \quad z_k = \sum_{j \in F_k} c_j x_j + \sum_{j \in S_k^1} c_j$$

$$\text{s.t.} \quad \sum_{j \in F_k} a_{ij} x_j \leq b_i - \sum_{j \in S_k^1} a_{ij} = s_i , \qquad (8.6)$$

$$i = 1,\ldots,m$$

$$x_j = 0,1 , \; j \in F_k$$

Suppose we substitute for the inequality constraints (8.6) the single constraint

$$\sum_{i=1}^{m} u_i \sum_{j \in F_k} a_{ij} x_j \leq \sum_{i=1}^{m} u_i s_i \qquad (8.7)$$

where $u = (u_1, u_2, \ldots, u_m) \geq 0$. Clearly, (8.7) is weaker than (8.6) since any set of x_j's satisfying the latter also satisfy the former but the converse need not be true, that is,

$S_k(u) \supseteq S_k$

where S_k and $S_k(u)$ are the set of binary feasible solutions to (8.6) and (8.7), respectively. Clearly, if $S_k(u) = \phi$, then $S_k = \phi$ also. To achieve the best surrogate constraint, we seek the multipliers u^* which yield

$$\min_{u \geq 0} \max_{x_j \in S_k(u)} \sum c_j x_j$$

$$x_j = 0,1$$

Let $g(u) = \max \sum_{j \in F_k} c_j x_j$; $x \in S_k(u)$ and let $g'(u)$ denote the maximum with the integrality restrictions on the x_j's dropped:

$$g'(u) = \text{maximum} \sum_{j \in F_k} c_j x_j$$

$$\text{s.t.} \sum_{i=1}^{m} u_i \sum_{j \in F_k} a_{ij} x_j \leq \sum_{i=1}^{m} u_i s_i$$

$$0 \leq x_j \leq 1 \ , \ j \in F_k$$

Then the best multipliers to this continuous version of Q_k provide an excellent estimate of the multipliers u^*. If we denote the best multipliers by u^0, then

$$g'(u^0) = \min_{u \geq 0} g'(u)$$

It is easy to demonstrate that u^0 is given by the optimal (dual) variables to the dual LP to Program Q_k

with the integrality requirements removed, that is, the optimal dual variables to the LP

$$\text{minimize} \quad \sum_{i=1}^{m} u_i s_i + \sum_{j \in F_k} w_j$$

$$\text{s.t.} \quad \sum_{i=1}^{m} a_{ij} u_i + w_j \geq c_j \quad \text{for} \quad j \in F_k$$

$$u_i \, , \, w_j \geq 0 \quad \text{for all} \quad i \, , \, j$$

Once the (best) surrogate constraint is generated, it can be used to generate bounds and to fathom nodes as explained above.

X. Introduce Relations Where None Existed. While constraints and relations complicate mathematical programming problems, in the sense of increasing their complexity and time to solution, constraints and relations may in fact be of great assistance in B&B methods because they limit the space of search. After all, considerations of infeasibility are among the most powerful methods in the B&B approach.

Consequently, one always seeks to establish constraints and relations among the unknown variables of the problem in order to reduce the space of search. Cuts and surrogate constraints, discussed above, are such devices. Here we wish to highlight the concept of generation of cuts that are based on derived relations from logical arguments, where none existed before. The reader may wish to view the (dominance) relations emphasized here as added examples of cuts, which indeed they are. However, there is a generic difference between the cuts discussed here and, say, those based on Benders' decomposition: these latter are based on mathematical programming arguments, while the former are based on permutational arguments without recourse to the mathematical programming model. Two applications illustrate what is meant here. The first is due to Emmons [8.18] and the second is due to Elmaghraby and Park [8.16]. We briefly review the latter.

Example 8.13. In their study of scheduling N jobs on M machines in parallel, Elmaghraby and Park [8.16] developed a set of dominance relations that helps reduce the search effort drastically. Their approach exemplifies the concept of dominance by inclusion, namely, a schedule with a particular property dominates all others that do not possess such property. (Below under Dictum XI we discuss an approach that exemplifies dominance by exclusion.) Suppose job j has processing time p_j and due date equal to its processing time. Furthermore, suppose there is a linear penalty of tardiness given by the product of π_j and $\max(0; T_j - p_j)$ for job j. It is desired to determine the schedule that minimizes the total penalty for tardiness. We use the following notation.

$Q_m = (m_1, m_2, \ldots, m_k)$ denotes a sequence of k jobs on machine m, $m = 1,2,\ldots,M$

Subscript m_j refers to the jth job on machine m. The subscript m_ℓ refers to the last job on machine m

s_i: the start time of job i. The start time of the last job on machine m is denoted by s_{m_ℓ}

T_i: the completion time of job i in a given sequence. The completion time of the last job on machine m is denoted by T_{m_ℓ}

$r_i = \pi_i / p_i$ and we assume that the jobs are numbered in nonincreasing ratio r_i, the so-called natural order

The dominance relations are presented in terms of properties of the optimal schedules or as precedence relations between pairs of jobs. (These latter assertions

can be easily translated into dominance relations.)

(i) For a schedule Q to be optimal, it is necessary for $Q_m = (m_1, m_2, \ldots, m_k)$ to be sequenced in the natural order.

(ii) There exists an optimal schedule in which job 1 is scheduled first.

(iii) For a schedule Q to be optimal it is necessary that

$$T_{i_\ell} \geq s_{j_\ell}$$

for every pair of machines i and j

(iv) Using the convention, job i precedes job j means that $s_i \leq s_j$, we have that if

$$p_i \leq p_j \quad \text{and} \quad \pi_i \geq \pi_j$$

then job $\underline{i}$ precedes job $\underline{j}$ in an optimal schedule.

(v) If $\pi_j = \pi$, a constant for all jobs, then there exists an optimal schedule in which jobs are assigned in their natural order from machine 1 to machine M in rotation.

(vi) Let $H = \sum_{i=1}^{N} p_i / M$, representing the processing interval if job splitting (including parallel operation on two or more machines) were permitted. Then in an optimal schedule on <u>two</u> machines, job $\underline{h}$ precedes job $\underline{k}$ if the following two conditions are satisfied.

(a) $r_h > r_{h+1}$ and $r_{k-1} > r_k$

$$(b) \quad \sum_{1}^{h-1} p_i \leq H - \frac{1}{2} \max_{i \in N} p_i - \sum_{k}^{n} p_i$$

As an example of translating the above statement into a dominance statement, consider (iii) above. Suppose we have a partial schedule of k jobs, denoted by $P(k)$, on three machines as shown by the solid lines in Figure 8.10. Then the completion of $P(k)$ which has job $(k+1)$ on machine 2 (the broken segment) dominates all other completions. In fact, if p_{k+1} (the processing time of job $k+1$) is such that $T_{2_\ell} + p_{k+1} < \min \left(T_{1_\ell}, T_{3_\ell} \right)$, which is the case represented in the figure, then the completion of the partial schedule $P(k+1)$ itself which has job $k+2$ placed on machine 2 dominates all other completions!

XI. Exclusion and Decomposition. Exclusion and decomposition constitute another form of derived relations based on permutational arguments. The only

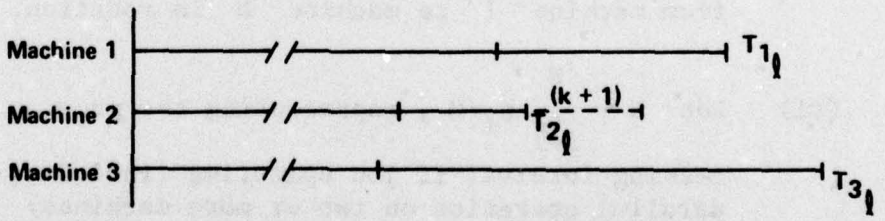


Figure 8.10 - A Gantt chart.

difference between exclusion cuts and the cuts discussed in Dictum X above is that here dominance is based on excluding a possibility. This is a weaker condition than the inclusion arguments discussed in X. On the other hand, decomposition arguments essentially "break up" the original space into subspaces with the links between them established optimally. Clearly, such decomposition reduces the burden of enumeration considerably. For instance, in a ten-job scheduling problem on a single facility the original space has $10! = 3,628,800$ points to be enumerated. If it can be established that optimality requires that a subset of four jobs must precede a subset of six jobs, then the enumeration is reduced to only $4! \times 6! = 17,280$ alternatives!

Example 8.14. An excellent illustration of such exclusion arguments and decomposition procedures was provided by Mitten and Tsou [8.41], referred to below as M&T, and whose paper was previously discussed in Example 8.9. They utilized the CBS condition (8.5) to establish some exclusion conditions as well as a condition for decomposition. Let $p = (p_1, p_2, \ldots, p_n)$ be a partial permutation of n elements $(n<N)$. Let $z \in \bar{S}_p$, where $\bar{S}_p = S - p$. The issue is to establish the conditions under which z is not a candidate for the $(n+1)$st position after p. Let σ_z be the set of elements of $\bar{S}_p$ whose R-value is not less than R_z at D_p:

$$\sigma_z = \{w \in \bar{S}_p : R_z(D_p) \leq R_w(D_p)\}$$

Clearly, $z \in \sigma_z$ and $\bar{S}_p - \sigma_z$ is the set of elements of $\bar{S}_p$ whose R-value is less than R_z at D_p. Let $A_z = \sum_{w\in\sigma_z} dw - \min_{w\in\sigma_z} dw$. Then A_z is a l.b. on the duration of σ_z beyond D_p and the following can be

established. If

(a) $\sigma_z \neq \bar{S}_p$, that is, if there exist elements $u \in \bar{S}_p$ for which $R_z(D_p) > R_u(D_p)$, and if

(b) $R_x(D) < R_y(D)$ for all $x \in \bar{S}_p - \sigma_z$, $y \in \sigma_z$, and all D s.t. $D_p \leq D \leq D_p + A_z$ (that is, the elements of $\bar{S}_p$ whose R-value is $< R_z$ retain that inequality for all time periods in the interval $[D_p, D_p + A_z]$)

then $(p,z,q) \notin P^*(0)$ for any permutation q of the remaining elements of the set $\bar{S}_p - \{z\}$. In particular, if $\sigma_z = \{z\}$ then z cannot follow p in position $n + 1$. The immediate consequence of this result is the formation of a preference table that gives for each value of D (there are only finitely many) the elements ranked in order of increasing $R_x(D)$. A check on conditions (a) and (b) is then facilitated, and would normally result in the elimination of candidates (branches in the B&B search) for position $n + 1$. To achieve decomposition, let $p \in P$ be a partial permutation and σ a nonempty subset of $\bar{S}_p$ to be partitioned into two nonempty subsets σ' and σ'' . As before, let $A = \sum_{w \in \sigma} d_w - \min_{w \in \sigma} d_w$, and $D' = \sum_{w \in \sigma} d_w$, Finally, suppose there exist two partial permutations $q^* \in P_{\sigma'}$ and $r^* \in P_{\sigma''}$ for which the following hold.

$C(q^*, D_p) \leq C(q, D_p)$ for all $q \in P_{\sigma'}$

$$C\left(r^*, D_p+D'\right) \leq C\left(r, D_p+D'\right) \quad \text{for all} \quad r \in P_{\sigma''},$$

$$R_x(0) < R_y(0) \quad \text{for all} \quad x \in \sigma' \,,\; y \in \sigma'' \quad \text{and all}$$

$$D \in [D_p, D_p+A_\sigma]$$

Then we conclude that $C[(p,q^*,r^*),0] \leq C[(p,u),0]$ for all $u \in P_{\bar{S}}$. In other words, the completion of p has been decomposed into two parts, q^* and r^*, with q^* definitely preceding r^*. Naturally, the check for such decomposition is guided by the (sufficient) condition of its realization. In particular for any given $\sigma \in \bar{S}_p$ one determines the two subsets σ' and σ'' based on the R-values; then one proceeds to seek the subsequences q^* and r^*. M&T report excellent computing results using the approaches outlined above.

8.6 Miscellaneous Dicta

In this final section we present three dicta.

XII. It Is Sometimes Better To Use a Not-So-Economical Model Because It Is More Amenable To Computing "Tricks". Geoffrion and Graves' treatment [8.28] of a multi-commodity distribution system provides an illustration. Example 8.15. The problem statement from [8.28] is as follows. There are several commodities produced at several plants with known production capacities. There is a known demand for each commodity at each of a number of customer zones. This demand is satisfied by shipping via regional distribution centers (DCs), with each customer zone being assigned exclusively to a single DC. There are lower as well as upper bounds on allowable total annual throughput of each DC. The possible locations for the DCs are given, but the particular sites to be used are to be selected so as to result in the least total distribution cost. The DC costs are expressed as fixed charges (imposed only on

the sites actually used) plus a linear variable charge which is a function of the "throughput" at the DC. Transportation costs are taken to be linear.

The mathematical model adopted was the following MLP.

$$\text{minimize}_{x \geq 0 \text{ and } y,z = 0,1} \quad \sum_{i,j,k,\ell} c_{ijk\ell}\, x_{ijk\ell} + \sum_{k} \left[f_k z_k + v_k \sum_{i,\ell} D_{i\ell}\, y_{k\ell} \right] \tag{8.8}$$

$$\text{s.t.} \quad \sum_{k,\ell} x_{ijk\ell} \leq S_{ij} \quad \text{all } i, j \tag{8.9}$$

$$\sum_{j} x_{ijk\ell} = D_{i\ell}\, y_{k\ell} \quad \text{all } i, k, \ell \tag{8.10}$$

$$\sum_{k} y_{k\ell} = 1 \quad \text{all } \ell \tag{8.11}$$

$$\underline{V}_k z_k \leq \sum_{i\ell} D_{i\ell}\, y_{k\ell} \leq \bar{V}_k z_k \quad \text{all } k \tag{8.12}$$

Linear configuration constraints on one or both of y and z (8.13)

The notation is explained as follows.

$x_{ijk\ell}$: the amount of commodity i shipped from plant j through DC k to customer zone ℓ

$c_{ijk\ell}$: average unit cost of production and shipping corresponding to $x_{ijk\ell}$

z_k : a 0,1 variable, equal to 1 if a DC is acquired at site k, and 0 otherwise

f_k : the fixed cost associated with acquiring the facility at site k

$y_{k\ell}$: a 0,1 variable, equal to 1 if DC k serves customer zone ℓ, and 0 otherwise

$D_{i\ell}$: demand for commodity i in zone ℓ

v_k : variable unit cost of throughput in a DC at site k

S_{ij} : production (supply) capacity for commodity i at plant j

$\underline{V}_k, \bar{V}_k$: minimum and maximum allowed total annual throughput for a DC at site k

It is not difficult to see that this model is too liberal in the number of equations, a condition usually avoided by workers in LP (continuous, integer or mixed). There is an obvious opportunity to economize on the number of constraints of type (8.10) without changing the logical content of the model in any way: replace (8.10) by

$$\sum_{j,k} x_{ijk\ell} = D_{i\ell} \quad \text{all } i, \ell \tag{8.10a}$$

$$\sum_{i,j} x_{ijk\ell} = \left(\sum_{i} D_{i\ell}\right) y_{k\ell} \quad \text{all } k, \ell \tag{8.10b}$$

This formulation handles the two functions of (8.10) separately, ensuring that all demands are met, and forcing the appropriate logical relationship between the x's and y's. But in any particular application it presents a saving in the number of constraints over (8.10). For instance, suppose $i = 100$, $j = 10$, $k = 20$, and $\ell = 1000$. Equations (8.10) would yield $(100)(20)(1000) = 2{,}000{,}000$ equations; while (8.10a) and (8.10b) would yield only $(100)(1000) + (20)(1000) = 120{,}000$ equations! However, the program of (8.8)-(8.13) is actually easier to solve than the program with (8.10a) and (8.10b) substituted for (8.10). The reason lies in the application of the Benders'

decomposition algorithm which yields excellent "cuts" and easy-to-solve transportation-type subproblems in the first case, and extremely weak "cuts" in the second case!

XIII. Do Not Optimize the Relaxed Problem; Just Obtain a Bound on Its Solution. In spite of the rather elementary nature of the concept, its implementation may spell the difference between an operational scheme and a difficult, albeit correct, one. The dictum is best illustrated by Eastman [8.9] and that of Little et al. [8.37] in the calculation of the bounds for the TSP. Both treatments approached the problem from the point of view of relaxing the tour constraint, which leaves the standard assignment problem. While Eastman solved the assignment problem to obtain its optimum, Little et al. were content with a ℓ.b. on its value. Clearly

$z^* \geq$ opt. value of assignment problem $\geq$ ℓ.b. on assignment problem

XIV. Preconditioning: Limit the Size of the Original Space To Be Searched. This is a manifestation of the well-known motto: "an ounce of prevention is worth a pound of cure." It draws attention to the need for the proper "conditioning" of the search space before the search is initiated. Two examples are provided in the literature. The first is due to Held, Karp and Sherashian [8.32] in the study of assembly line balancing, and the second is due to Elmaghraby [8.11] in a scheduling problem. We briefly discuss this latter.

Consider the scheduling of N jobs on a single facility, in which each job j has a processing time p_j, due date d_j, and a penalty function $\pi_j \max(0;T_j-d_j)$ which penalizes tardiness, where T_j is the completion time of job j deduced from the schedule. It is desired to determine the schedule that minimizes the total cost of tardiness. At the outset there are $N!$ different sequences, and as many points to be enumerated. However, through developing a dynamic programming model of the problem, the size of the search space was reduced to only 2^N. Branch-and-Bound was

then applied to this smaller space, with excellent results.

References

[8.1] Agin, N. (1966). Optimum seeking with branch and bound. Management Sci. 13 B176-B185.

[8.2] Ashour, S. (1972). Sequencing Theory. Lecture Notes in Economics and Mathematical Systems 69. Springer.

[8.3] Ashour, S., and M. N. Quraishi (1969). Investigation of various bounding procedures for production scheduling problems. International Journal of Production Research 7 249-252.

[8.4] Beale, E. M. L. (1965). A survey of integer programming. Operational Res. Quart. 16 219-229.

[8.5] Benders, J. F. (1962). Partitioning procedures for solving mixed-variable programming problems. Numer. Math. 4 238-252.

[8.6] Breu, R., and C. A. Burdet (1973). Branch and bound experiments in 0-1 programming. Management Science Research Group Report, Carnegie-Mellon University (August).

[8.7] Clark, C. E. (1961). The optimum allocation of resources among the activities of a network. Journal of Industrial Engineering 12 11-17.

[8.8] Davis, E. W. (1973). Project scheduling under resource constraints--historical review and categorization of procedures. American Institute of Industrial Engineering Transactions 5 257-313.

[8.9] Eastman, W. L. (1968). A solution to the traveling salesman problem. Paper presented at the Summer Meeting of the Econometric Society, Cambridge, Massachusetts (August).

[8.10] Eastman, W. L., S. Even, and J. M. Isaacs

(1964). Bounds for optimal scheduling of jobs. Management Sci. 11 268-279.

[8.11] Elmaghraby, S. E. (1968a). The one machine sequencing problem with delay costs. Journal of Industrial Engineering 19 105-108.

[8.12] Elmaghraby, S. E. (1968b). The determination of optimal activity duration in project scheduling. Journal of Industrial Engineering 19 48-51.

[8.13] Elmaghraby, S. E. (1968c). The sequencing of n jobs on m parallel processors with extensions to scarce resource problem of activity networks. in Proceedings of the Inaugural Conference of the Scientific Computation Center and the Institute of Statistical Studies and Research, Cairo, Egypt, 230-255.

[8.14] Elmaghraby, S. E., and A. K. Mallik (1973). The scheduling of a multi-product facility. in S. E. Elmaghraby (ed.) Symposium on the Theory of Scheduling and its Applications. Lecture Notes in Economics and Mathematical Systems 86. Springer. 244-277.

[8.15] Elmaghraby, S. E., and L. P. Dix (1973). Scheduling a single facility under constant demand and fixed production rate. OR Report No. 87, North Carolina State University (May).

[8.16] Elmaghraby, S. E., and S. Park (1974). Scheduling jobs on a number of identical machines. American Institute of Industrial Engineering Transactions 6 1-13.

[8.17] Elshafei, A. N. (1974). On solving the capacitated facility location problem with concave cost functions. OR Reports Nos. 92, 93, North Carolina State University (June).

[8.18] Emmons, H. (1969). One machine sequencing to minimize certain functions of job tardiness. Operations Res. 17 701-715.

[8.19] Faaland, B. H. (1972). Estimates and bounds

on computational effort in the accelerated bound-and-scan algorithm. Technical Report No. 35, Department of Operations Research, Stanford University (May).

[8.20] Faaland, B. H., and F. S. Hillier (1972). The accelerated bound-and-scan algorithm for integer programming. Technical Report No. 34, Department of Operations Research, Stanford University (May).

[8.21] Falk, J. E., and J. L. Horowitz (1972). Critical path problems with concave cost-time curves. Management Sci. 19 446-455.

[8.22] Falk, J. E., and R. M. Soland (1969). An algorithm for non-separable non-convex programming problems. Management Sci. 15 550-569.

[8.23] Forrest, J. J. H., J. P. H. Hirst, and J. A. Tomlin (1974). Practical solution of large mixed integer programming problems with UMPIRE. Management Sci. 20 736-773.

[8.24] Fulkerson, D. R. (1961). A network flow computation for project cost curve. Management Sci. 7 167-178.

[8.25] Garfinkel, R. S., and G. L. Nemhauser (1972). Integer Programming. Wiley.

[8.26] Geoffrion, A. M. (1967). Integer programming by implicit enumeration and Balas method. SIAM Rev. 9 178-190.

[8.27] Geoffrion, A. M. (1970). Elements of large scale mathematical programming. Parts I and II. Management Sci. 16 652-691.

[8.28] Geoffrion, A. M., and G. W. Graves (1974). Multicommodity distribution system design by Benders decomposition. Management Sci. 20 822-844.

[8.29] Glover, F. (1965). A multi-phase dual algorithm for the zero-one integer programming problem. Operations Res. 13 879-919.

[8.30] Greenberg, H., and R. L. Hegerick (1970). A branch and bound algorithm for the knapsack problem. Management Sci. 16 327-332.

[8.31] Held, M., and R. M. Karp (1970). The traveling salesman problem and minimum spanning trees. Operations Res. 18 1138-1162.

[8.32] Held, M., R. M. Karp, and R. Sherashian (1963). Assembly line balance--dynamic programming with precedence constraints. Operations Res. 11 442-459.

[8.33] Ignall, E., and L. Schrage (1965). Application of the branch-and-bound technique to some flow shop scheduling problems. Operations Res. 13 400-412.

[8.34] Kan, A. H. G. Rinnooy (1973). The machine scheduling problem. The Mathematical Center, Amsterdam (August).

[8.35] Land, A. H., and A. G. Doig (1960). An automatic method for solving discrete programming problems. Econometrica 20 497-520.

[8.36] Lawler, E. L., and D. E. Wood (1966). B&B methods: a survey. Operations Res. 14 699-717.

[8.37] Little, J. D. C., K. G. Murty, D. W. Sweeney, and C. Karel (1963). An Algorithm for the traveling salesman problem. Operations Res. 11 972-989.

[8.38] Lomnicki, Z. A. (1965). A branch-and-bound algorithm for the exact solution of the three machine scheduling problem. Operational Res. Quart. 16 89-100.

[8.39] Mitten, L. G. (1970). Branch-and-bound methods: general formulation and properties. Operations Res. 18 24-34.

[8.40] Mitten, L. G. (1973). Branch and bound: a general formulation and survey of the literature. Research Memorandum, University of British Columbia (August).

[8.41] Mitten, L. G., and C. A. Tsou (1973). Efficient solution procedures for certain scheduling and sequencing problems. in S. E. Elmaghraby (ed.) Symposium on the Theory of Scheduling and Its Applications. Lecture Notes in Economics and Mathematical Systems 86. Springer 127-142.

[8.42] Mitten, L. G., and A. R. Warburton (1973). Implicit enumeration procedures. Research Memorandum, University of British Columbia (March).

[8.43] Rech, P., and L. G. Barton (1970). A non-convex transportation algorithm. in E. M. L. Beale (ed.) Applications of Mathematical Programming Techniques. American Elsevier. 250-260.

[8.44] Srinivasan, V., and G. L. Thompson (1973). Solving scheduling problems by applying cost operations to assignment models. in S. E. Elmaghraby (ed.) Symposium on the Theory of Scheduling and its Applications. Lecture Notes in Economics and Mathematical Systems 86. Springer 399-425.

Chapter 9

RECENT ANALYTICAL ADVANCES IN FACILITY LAYOUT AND LOCATION: A SURVEY*

Richard L. Francis
University of Florida

9.1 Introduction

Due to a number of relatively recent bibliographies [9.8], [9.15], surveys [9.6], [9.10], and books [9.9], dealing with location literature, we choose here to restrict the discussion to recent analytical advances in solving layout problems, and selected location problems that may also be viewed as layout problems.

The basic distinction we make between location problems and layout problems is that with location problems the activities or facilities to be located may be idealized as points, whereas this is not the case with layout problems, where the facilities to be laid out take up a positive area, and one often wishes either to determine, or to prespecify, the shape of each facility. Layout problems are quite diverse and include, as examples, laying out files on a computer tape, electronic modules on a circuit board, instruments on an aircraft cockpit panel, machines within a shop, furniture for one's spouse, departments within a plant, offices within a building, or items within a warehouse. Further, layout problems are closely related to such problems as urban redistricting [9.2], [9.14], regional partitioning, and political redistricting [9.21].

In spite of the variety and scope of layout problems, they have received relatively little attention when compared to location problems. Reasons for this relative lack of attention seem to include the following. (a) Layout problems are often of a smaller scope than location problems. (b) It may be difficult to

*Research by the author for his publications cited in the references was partially supported by the Army Research Office - Durham under Contract DAO C04 68C 002 and by the National Science Foundation under Grant GK-41224.

measure the value of a layout (rearranging furniture being an example). (c) It may be difficult to agree upon the class of facility shapes to choose from. (d) Solutions to layout problems are often easy to criticize after the fact. (e) One of the basic underlying analytical problems is known to be at least as difficult as the travelling salesman problem. In spite of the difficulties involved in studying layout problems, it seems clear that they deserve more analytical attention due to the pervasiveness of such problems, the growing body of optimization theory and location theory knowledge that can be drawn upon in solving layout problems, the evident opportunities to improve upon means commonly used in practice (for example, three-dimensional models which are quite descriptive but also difficult to use to evaluate many alternatives), and the analytical challenge of these problems.

9.2 Problem Classifications, Statements, and Results

Tractable objective functions for layout problems are typically of one of two classes; for convenience we designate problems either as Class I or II, depending upon the type of objective function of the problem. For Class I problems, costs are a function of distance, such as average or maximum distance, between activities to be located and points at known locations. An example of a Class I problem would be a warehouse layout problem where items (activities) move directly in and out of storage via docks at known locations. For Class II problems, costs are a function of distance between activities being located. An example of a Class II problem would be a machine shop layout problem where items enter the shop and may travel to and from a number of machines prior to leaving the shop. In either of the two problem classes, the typical objective is to find a layout minimizing the cost of item travel. Clearly a real layout problem can be both a Class I and a Class II problem; however, for purposes of exposition, little is lost by considering only individual Class I and Class II problems.

Methods for solving Class I problems are relatively well-known. Often closed form answers can be obtained; otherwise, discrete approximation problems can be constructed and solved using any one of a variety of

network flow algorithms. Such problems are discussed extensively in [9.9], and will receive no further attention here.

A large collection of Class II problems is closely related to a problem commonly called the quadratic assignment problem for which substantial literature discussions are given in [9.11] and [9.20]. Since the problem is so basic, we state one version of the problem as follows: assume m activities are to be assigned to m sites, with all activities and sites being the same size, so that a permutation $\sigma = (\sigma(1), \ldots,\sigma(m))$ of the integers $1,2,\ldots,m$ may represent an assignment, where $\sigma(i)$ is the number of the site to which activity i is assigned. Let $d(i,j)$ represent an appropriately determined distance from site i to site j, and assume a "weight" matrix $W = (w_{ij})$ is given, where w_{ij} represents a known cost per unit distance incurred, for a given time period, due to "travel" between activity i and activity j. The total cost of an assignment, denoted by $TC(\sigma)$, is then given by

$$TC(\sigma) = \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij}\, d(\sigma(i),\sigma(j))$$

and the problem of interest is to find an assignment of activities to sites of least total cost. This problem is known to be at least as difficult as the travelling salesman problem [9.22], so that heuristic procedures, or branch-and-bound procedures, are commonly used to solve it [9.11].

When more is assumed about the quadratic assignment problem, there are better ways than heuristics, or branch-and-bound, to solve the problem. The assumptions all require the sites to be representable by points along the line, so that the resulting problem is rather idealized. The analyses of the resulting problems do have some direct applications, however, such as locating activities along a conveyor, aisle, or street, and are also of interest because they provide some qualitative insight into the planar location problems. When the sites are m distinct adjacent integers, Karp and Held [9.12] have shown that the problem, which they call the

module placement problem, may be formulated as a dynamic programming problem. Adolphson and Hu [9.1] have recently related the module placement problem to the theory of network flows; they represent the matrix W as a graph G with m nodes, where w_{ij} is associated with the arc connecting nodes i and j, and give, for the case when G is a tree, an algorithm which finds the global optimum solution in $O(m \log m)$ steps. Simmons [9.23], [9.24] develops a branch-and-bound algorithm for solving a version of the module placement problem for which activities are not the same size. The algorithm appears computationally useful only for $m \leq 9$.

Next we formulate and state results for two problems closely related to the module placement problem. Suppose positive parameters u_i and A_i are given for activity i, $1 \leq i \leq m$, with u_i a measure of the relative use of activity i (for example, the percent of items per week moving to and from machine i), and with A_i being the length of the interval taken up by activity i. The first problem is the following.

Problem PMS

$$\text{minimize} \quad \sum_{1 \leq i < j \leq m} [w(u_i,u_j)|x_i - x_j|]$$

$$\text{s.t.} \quad |x_i - x_j| \geq c(A_i,A_j) \quad \text{for} \quad 1 \leq i \leq j \leq m$$

In the statement of this problem, x_i is the midpoint of the interval taken up by activity i, $c(A_i,A_j)$ is a specified lower bound on the distance between the midpoints of activities i and j, and $w(u_i,u_j)$ is a specified nonnegative function of u_i and u_j.
The second problem is obtained by changing the objective function to obtain

Problem PMM

$$\text{minimize} \quad \underset{1\le i<j\le m}{\text{maximum}} \left[w(u_i,u_j)\,|x_i-x_j|\right]$$

$$\text{s.t.} \quad |x_i-x_j| \ge c(A_i,A_j) \quad \text{for} \quad 1 \le i \le j \le m$$

Note that PMS is equivalent to the module placement problem when $c(A_i,A_j) = 1/2$, $1 \le i < j \le m$, and $w(u_i,u_j) = w_{ij}$. Problem PMM is a minimax analog of PMS. Let the activities be numbered so that $(u_i/A_i) \ge (u_{i+1}/A_{i+1})$, $1 \le i \le m-1$. If the items are located so that

$$x_m \le x_{m-2} \le x_{m-4} \le \cdots \le x_5 \le x_3 \le x_1 \le x_2 \le x_4 \le \cdots \le x_{m-3} \le x_{m-1} \tag{9.1}$$

(or so that all inequalities in (9.1) are reversed) when m is an odd integer, and located so that

$$x_{m-1} \le x_{m-3} \le \cdots \le x_5 \le x_3 \le x_1 \le x_2 \le x_4 \le \cdots \le x_{m-2} \le x_m \tag{9.2}$$

(or so that all inequalities in (9.2) are reversed) when m is an even integer, and adjacent activities are located as near one another as possible subject to satisfying the interdistance constraint, we say that the activities are arranged in <u>alternating order</u>. Pratt [9.22] has shown that an alternating order arrangement is optimal to the version of PMS for which $w(u_i,u_j) = u_i u_j$. Chan and Francis [9.4] have shown than an alternating order arrangement is optimal to a number of versions of PMS and PMM now to be specified. In particular, define three special cases of PMS and PMM as shown in Table 9.1.

For a number of the cases in Table 9.1, arrangements other than alternating order are also optimal; details are given in [9.4]. Note that the alternating order solutions are, roughly speaking, all either such

Table 9.1 Special Cases of Problems PMS and PMM

Special Case	Form of $w(u_i,u_j)$	Form of $c(A_i,A_j)$
PMS 1	$f(u_i) + f(u_j)$	unity
PMS 2	$f(u_i)f(u_j)$	unity
PMS 3	unity	$(A_i+A_j)/2$
PMM 1	nondecreasing in each argument	unity
PMM 2	unity	$(A_i+A_j)/2$

that activities with largest usage (u_i), or smallest size (A_i) are closest, and in the middle; this is the major qualitative conclusion obtained from the analysis of PMS and PMM, and suggests similar approaches for solving analogous planar layout problems. When neither of the functions $w(.,.)$ nor $c(.,.)$ is unity, problems PMS and PMM become much more difficult to solve; whether or not any useful qualitative insights are obtainable remains to be determined.

Next, a planar layout problem is considered. Consider unit squares in the plane, each having a lattice point (a point with integer entries) as a center and sides parallel to the axes, as representing potential sites for activities. Any choice of m distinct lattice points, which specifies a choice of m distinct unit squares, is called a configuration of size m, and denoted by C_m. A configuration of size m may be considered either to represent a choice of m sites for activities, or alternatively, the unit squares may be considered modular construction units constituting a single facility. Given any two points $X_1 = (x_1,y_1)$ and $X_2 = (x_2,y_2)$ in the plane, the rectilinear distance between X_1 and X_2 will be denoted by $r(X_1,X_2)$, where by definition,

$r(X_1,X_2) = |x_1-x_2| + |y_1-y_2|$

The use of the rectilinear distance is often appropriate in commonly occurring situations where travel is carried out on a set of rectilinear aisles. Thus, for a configuration of size m, the rectilinear distance between any two lattice points may be considered to be an approximation to the distance travelled between the centers of the two sites represented by unit squares. For any configuration C_m, the diameter of C_m, denoted by $d(C_m)$, is defined to be the maximum rectilinear distance between all pairs of lattice points in C_m; $d(C_m)$ is always a positive integer, and may be considered to represent the distance between any two sites in a configuration that are farthest apart, in the rectilinear sense. The problem of finding a configuration of size m of minimum diameter may then be considered to be a minimax facility configuration problem. To state the main results for this problem, let $f(m)$ be the minimum diameter of all configurations of size m, and define the function $g(i)$, which is readily verified to be integer-valued and strictly increasing, as follows.

$g(i) = (i^2+2i+2)/2$ if i is an even, nonnegative integer

$= (i^2+2i+1)/2$ if i is an odd, positive integer

Then it can be proven that, for any positive integer i, $f(m) = i$ for every integer m satisfying $g(i-1) + 1 \leq m \leq g(i)$. Further, a configuration C_m^* is a minimax configuration if and only if $d(C_m^*) = i$, where

$i = \min \{j: m \leq g(j), j \text{ a positive integer}\}$

Proofs of these results, as well as a simple geometric procedure for constructing minimax configurations, may

be found in [9.7] or [9.9].

To this point the discussion has dealt with the case where the activities to be located are indivisible, in the sense that an activity is assigned to a specific site, or is represented by an interval on the line. We now consider situations where activities may be divisible; a specific example of such a situation would be a building served by a parking lot, with the two halves of the parking lot on opposite sides of a building. In order to pursue the discussion, a more general definition of a layout is needed. Suppose that the vector $A = (A_1, A_2, \ldots, A_m)$ consisting of m positive entries is given and that L is a region in E^n, Euclidean n-space, of measure at least $A_1 + \ldots + A_m$. Define $\{S_1, \ldots, S_m\}$ to be a layout, where S_i is a compact subset of L, of measure A_i, and S_i and S_j share no common interior points, for $1 \leq i \neq j \leq m$. As an example of the definition, $\{S_1, \ldots, S_m\}$ might represent a warehouse layout, if m items are to be laid out in a warehouse, with L representing the floor of the warehouse, and S_i (contained in L) representing the region of the floor taken up by item i, with A_i being the area of S_i, for $1 \leq i \leq m$. As a second example, $\{S_1, S_2\}$ might represent the layout of a building and adjacent parking lot, with S_1 representing the region taken up by the parking lot, and S_2 the region taken up by the building.

With the foregoing definition of a layout, two closely related minimax layout problems, and their solutions, may be stated. For convenience, define

$$B = A_1 + A_2 + \cdots + A_m$$

$$B_p = B - A_p \quad \text{for} \quad 1 \leq p \leq m$$

$$B_{1,2} = B - A_1 - A_2$$

and let activities be numbered so that

$$A_1 \geq A_2 \geq \max (A_i: \ 2 \leq i \leq m) \tag{9.3}$$

Given any layout $\{S_1,\ldots,S_m\}$ define

$$F_{ij}(S_i,S_j) = \max \{d(x_i,x_j): \ x_i \in S_i, x_j \in S_j\}$$

where $d(x_i,x_j)$ is either the rectilinear distance when $L = E^2$, or the absolute value distance when $L = E^1$; then define an objective function for the layout by

$$F(S_1,\ldots,S_m) = \max \{F_{ij}(S_i,S_j): \ 1 \leq i < j \leq m\} \tag{9.4}$$

The minimax layout problem is to find a layout that minimizes (9.4); roughly speaking, such a layout has the property that it minimizes, from among all layouts, the greatest distance between any two activities. When $n = 1$, so that the layout is a layout on the line, Papineau and Francis [9.18] show that a layout $\{S_1^*,\ldots,S_m^*\}$ minimizes (9.4) if and only if, for some $1 \leq p \leq m$ such that $A_1 = A_p$, $T_p^* \equiv \cup\{S_i^*: \ 1 \leq i \leq m, i \neq p\}$ is a closed interval of length B_p, and $S_p^* = S_p'^* \cup S_p'^{**}$, where $S_p'^*$ and $S_p''^*$ are both closed intervals of length $A_p/2$ abutting opposite ends of T_p^*; further, $F(S_1^*,\ldots,S_m^*) = B - (A_p/2)$. The answer has some intuitive appeal, since if the "largest" activity (S_p^*) were not placed on the "ends" of the layout, the distance between other activities would necessarily be increased.

When $n = 2$, so that the layout is a planar layout, Papineau et al. [9.19] have shown that an attainable lower bound on (9.4) is given by $\sqrt{2}\,\alpha_0$, with

$\alpha_0 \equiv \min(\alpha_I, \alpha_{II})$, where

$\alpha_I \equiv (1/2)\left[B^{1/2} + B_1^{1/2}\right]$

$\alpha_{II} \equiv (1/2)\left[(A_1)^{1/2} + (2B_1)^{1/2}\right]$, $B_{1,2} < (A_1B_1/2)^{1/2} < B_1$

$\equiv (1/2)\left[(B_{1,2} + B_1)(B_{1,2} + B_2)/B_{1,2}\right]^{1/2}$,

$(A_1B_1/2)^{1/2} \leq B_{1,2}$

$\equiv \infty$, otherwise

When $\alpha_0 = \alpha_I$, any layout $\{S_1^*, \ldots, S_m^*\}$ for which

$S_1^* \equiv \{(x,y): (B_1/2)^{1/2} \leq |x| + |y| \leq (B/2)^{1/2}\}$

$T_2^* \equiv \cup\{S_i^*: 2 \leq i \leq m\} = \{(x,y): |x| + |y| \leq (B_1/2)^{1/2}\}$

is a minimax layout. We note that T_2^* is a square of area B_1 with center at the origin, and that $\cup\{S_i^*: 1 \leq 1 \leq m\}$ is also a square of area B , with center at the origin; each edge of each square makes an angle of +45° or -45° with the x-axis. Further, the region taken up by activity 1 , which has the largest area, "surrounds" the other regions. In order to describe minimax layouts obtained when $\alpha_0 = \alpha_{II}$, imagine a cross composed of two rectangles making right angles with each other, let R_1 and R_2 denote the horizontal and vertical rectangles respectively, let $S_1'^*$,

and $S_1^{''*}$ represent that part of R_1 to the left and right respectively of R_2, let $S_2^{'*}$ and $S_2^{''*}$ represent that part of R_2 above and below R_1 respectively, and let $T_{1,2}^{*} \cup S_2^{'''*}$ represent $R_1 \cap R_2$. Let a and d be the width and height respectively of both $S_1^{'*}$ and $S_1^{''*}$, and let b and c be the width and height respectively of both $S_2^{'*}$ and $S_2^{''*}$, so that the width and height of $T_{1,2}^{*} \cup S_2^{'''}$ are given by b and d respectively. The symbols a, b, c and d will be called the dimensions of the layout, and each will have a subscript 1 or 2 attached to indicate whether one type or another type of a layout is being considered. When $\alpha_0 = \alpha_{II}$ and $B_{1,2} < (A_1B_1/2)^{1/2}$ (after the cross is rotated clockwise 45° from the vertical, so that each edge of each rectangle makes an angle of ±45° with the x-axis), then $\{S_1^*,\ldots,S_m^*\}$ is a minimax layout if it has dimensions

$$2a_1 = d_1 = (A_1)^{1/2},\ b_1 = (1/2)(2B_1)^{1/2},$$

$$c_1 = (1/2)\left[(2B_1)^{1/2} - (A_1)^{1/2}\right]$$

and is defined by

$$S_1^* = S_1^{'*} \cup S_1^{''*}$$

$$S_2^* = S_2^{'*} \cup S_2^{''*} \cup S_2^{'''*}$$

$$T_{1,2}^{*} = \cup\{S_i^* : 3 \leq i \leq m\}$$

$S_2^{'''*}$ may constitute any portion of $R_1 \cap R_2$ of area $A_2 - 2b_1c_1$. When $\alpha_0 = \alpha_{II}$ and $(A_1B_1/2)^{1/2} \leq B_{1,2}$ (after the cross is rotated clockwise 45° from the vertical) then $\{S_1,\ldots,S_m^*\}$ is a minimax layout if it has the dimensions

$$a_2 = (A_1/2)[(B_{1,2}+B_1)/(B_{1,2}+B_2)B_{1,2}]^{1/2}$$

$$b_2 = [B_{1,2}(B_{1,2}+B_1)/(B_{1,2}+B_2)]^{1/2}$$

$$c_2 = (A_2/2)[(B_{1,2}+B_2)/(B_{1,2}+B_1)B_{1,2}]^{1/2}$$

$$d_2 = [B_{1,2}(B_{1,2}+B_2)/(B_{1,2}+B_1)]^{1/2}$$

and is defined by $S_1^* = S_1^{'*} \cup S_1^{''*}$, $S_2^* = S_2^{'*} \cup S_2^{''*}$, $S_2^{'''*} = \phi$, and $T_{1,2} = \cup\{S_i^*: 3 \leq i \leq m\}$. These two "cross" layouts are clearly unorthodox, and should perhaps best be viewed as design benchmarks, in the sense that they provide an absolute basis of comparison for implementable layouts. Also, they provide a basis for further analytical enrichment, and do have an intuitively appealing property in that the two regions taken up by activities having first and second largest areas (recall (9.3)) are not enclosed by other regions.

It should be clear that a variety of Class II problems remain unsolved; we mention a few unsolved problems with the hope of encouraging others to study such problems. The first unsolved problem involves finding the shapes and relative locations of two or more planar sets of known area to minimize the total weighted average distance among the sets; the closest version of this problem which is solved is due to Newman [9.17] for two

sets which are subsets of the line instead of the plane. A second problem would be to impose the additional restriction upon the first problem that all sets must be rectangles. A third problem would be to add perimeter costs to either the first or second problem. A fourth and much more general problem would be to construct meaningful, analytically tractable objective functions that incorporate aspects besides distance. Hopefully it is now evident from this discussion that a great deal remains to be done in developing a cumulative, analytical body of knowledge for Class II problems. Class II problems are of course being "solved" in practice every day, but there is still much to be done in providing a theoretical basis to support (or not to support) the contention that they are being well solved.

References

[9.1] Adolphson, D., and T. C. Hu (1973). Optimal linear ordering. Paper presented at the Mathematical Programming Symposium, Stanford University.

[9.2] Carter, G. M., J. M. Chaiken, and E. Ignall (1972). Response areas for two emergency units. Operations Res. 20 571-594.

[9.3] Chan, A. W. (1974). On some facility location and layout problems. PhD dissertation, University of Florida.

[9.4] Chan, A. W., and R. L. Francis (1974). Some layout problems on the line with interdistance constraints and costs. Report UFLA-ISE-74-2, Department of Industrial and Systems Engineering, University of Florida.

[9.5] Elmaghraby, S. E. (1968). The sequencing of "related" jobs. Naval Res. Logist. Quart. 15 23-32.

[9.6] Elshafei, A. N., and K. B. Haley (1974). Facilities location: some formulations, methods of solution, applications, and computational experience. OR Report No. 90, North Carolina State University at Raleigh.

[9.7] Francis, R. L. (1973). A minimax facility-configuration problem involving lattice points. Operations Res. 21 101-111.

[9.8] Francis, R. L., and J. M. Goldstein (1974). Location theory: a selective bibliography. Operations Res. 22 400-410.

[9.9] Francis, R. L., and J. A. White (1974). Facility Layout and Location: An Analytical Approach. Prentice-Hall.

[9.10] Geoffrion, A. M. (1974). A guide to computer-assisted methods for distribution systems planning. Working Paper No. 216, Western Management Science Institute, University of California, Los Angeles. (June).

[9.11] Hanan, M., and J. M. Kurtzberg (1972). A review of the placement and quadratic assignment problems. SIAM Rev. 14 324-342.

[9.12] Karp, R. M., and M. Held (1967). Finite-state processes and dynamic programming. SIAM J. Appl. Math. 15 693-718.

[9.13] Larson, R. C. (1972). Urban Police Patrol Analysis. Massachusetts Institute of Technology Press.

[9.14] Larson, R. C., and K. A. Stevenson (1972). On insensitivities in urban redistricting and facility location. Operations Res. 20 595-612.

[9.15] Lea, A. C. (1973). Location-allocation system: an annotated bibliography. Discussion Paper No. 13, Department of Geography, University of Toronto.

[9.16] Neghabat, F. (1974). An efficient equipment-layout algorithm. Operations Res. 22 622-628.

[9.17] Newman, D. J. (1964). A parking lot design. SIAM Rev. 6 62-66.

[9.18] Papineau, R. L., and R. L. Francis (1974). A

minimax layout problem on the line involving distances between classes of objects. American Institute of Industrial Engineering Transactions 6 252-256.

[9.19] Papineau, R. L., R. L. Francis, and J. J. Bartholdi (1975). A minimax facility layout problem involving distances between and within facilities. American Institute of Industrial Engineering Transactions 7 (4).

[9.20] Pierce, J. F., and W. B. Crowston (1971). Tree-search algorithms for quadratic assignment problems. Naval Res. Logist. Quart. 18 1-36.

[9.21] Pollack, S. (ed.) (1972). Algorithmic approaches to political redistricting. College of Engineering, Institute of Public Policy Studies, University of Michigan.

[9.22] Pratt, V. R. (1972). An N log N algorithm to distribute N records in a sequential access file. in R. E. Miller, J. W. Thatcher, and J. D. Bohlinger (eds.) Complexity of Computer Computations. Plenum. 111-118.

[9.23] Simmons, D. M. (1969). One-dimensional space allocation: an ordering algorithm. Operations Res. 17 812-826.

[9.24] Simmons, D. M. (1971). A further note on one-dimensional space allocation. Operations Res. 19 249.

Part IV

PROBABILISTIC AND STATISTICAL MODELS

Chapter 10

REVIEW OF STATISTICAL PROBLEMS AND METHODS IN LOGISTICS RESEARCH*

S. Zacks
Case Western Reserve University and
The George Washington University

10.1 Introduction

This chapter provides a review of some problem areas in logistics research in which specific statistical methods have been developed. Statistical estimation procedures are commonly applied whenever operational analysts study stochastic systems with unknown or incomplete information on the distributions of the random variables under consideration. Certain areas of logistics research have revealed specific types of statistical problems. The present paper discusses the specific statistical problems in five of these areas. More specifically, we review here some developments that have taken place in the areas of demand prediction, adaptive inventory control, operational readiness, detection of wearout and surveillance. There are hundreds of papers in the various journals that may be ascribed or related to the five problem areas mentioned above. The present discussion is limited, however, to a relatively small number of papers that are specifically related to the research which the author has carried out for the Program in Logistics.

10.2 Demand Prediction

The statistical analysis of demand for spare parts in military systems occupies a central place in logistics research. These demand prediction studies are oriented towards special problems of inventory control, maintenance and replacement, surveillance, and so on. Many studies were performed on the demand patterns for spare parts in different types of military systems, aircraft,

*Preparation of this chapter was supported by the Office of Naval Research under Contract N00014-67-A-0214-0001 with the George Washington University.

ships, missiles, and others. A complete reference list of papers written on the subject of demand prediction will include dozens of papers. We refer the reader in particular to Reference [10.1].

The statistical methods for forecasting and predicting the demand for spare parts can be classified into two main types: (i) methods appropriate for extremely low demand patterns, and (ii) methods appropriate for regular and high rates of demand. The first type of extremely low demand has been observed especially in naval systems such as the Polaris system. This kind of low demand is apparently typical of a high percentage of spare parts in vessels. On the other hand, demand prediction for aircraft systems, fixed-wing or rotary, falls apparently in the second category. In the present exposition we focus attention on the methodological problems associated with these two types of demand patterns.

10.2.1 Statistical Methods for Low Demand. The study of demand patterns for spare parts for naval systems has been conducted at the Logistics Research Project of The George Washington University since its beginning. The first major study of demand behavior was completed in 1957. As summarized by Solomon [10.27] the context of the study was the so-called "allowance list problem." The data consisted of mechanical and electrical parts used by twelve submarines over a four-year period. The important conclusion was as follows.

> "The demand for items was extremely low and sporadic. About 75% of the items were not demanded at all. Moreover, during the entire four-year period for each submarine and its supply activity, 70% of the items demanded were demanded only once. Approximately 90% of the items demanded were demanded at most twice." [Furthermore] "almost all items that were demanded in one year were not demanded in another year."

These results were further reinforced by subsequent studies on the Polaris system (see [10.4]).

Haber, Sitgreaves, and Solomon [10.12] proposed a new methodology for estimating usage estimates of line

items. According to this methodology the prediction for the demand of parts is determined by the parent component of the system in which the part is installed. As usual, they distinguish between the range of parts in a component, that is, how many different parts are installed, and the depth of demand which is the frequency of demand for the individual part. Since many have not shown any demand, the authors of [10.12] adopted a binomial model concerning the number of parts in each component that show (at least one) demand. Accordingly, if the range of parts in a given component is N the model assumes that the parts fail independently and the probability for a demand θ is the same for all the parts in a given component, that is, the number of parts X showing demand is binomially distributed $B(N,\theta)$. Furthermore, since N is large and θ very small, this binomial distribution is approximated by a Poisson distribution with mean $\lambda = N\theta$. A structural model is proposed, which takes into account both the "component unreliability" and the "patrol severity" factors. Accordingly, it is assumed that the expected number of parts showing demand in the νth component at the ith patrol is $\lambda_{\nu i} = N_\nu C_\nu S_i$, where C_ν is the unreliability factor of the νth component and s_i is the severity factor of the ith patrol. It was also assumed that $\Sigma S_i = s$, where s is the number of observed patrols. Accordingly, the mean of the patrol severity factors is assumed to be 1. Haber, Sitgreaves, and Solomon proposed in [10.12] a simple method for estimating the factor C_ν and S_i from all the data. If we denote by $X_{\nu i}$ the number of parts showing demand in the νth component at the ith patrol then the proposed estimators are

$$\hat{S}_i = s \frac{\sum_\nu X_{\nu i}}{\sum_i \sum_\nu X_{\nu i}} \qquad (10.1)$$

and

$$\hat{c}_\nu = \sum_i X_{\nu i}/sN_\nu \qquad (10.2)$$

There is only a heuristic discussion in [10.12] leading to the construction of these estimators. An extension of the multiplicative model is given there for the purpose of estimating also specific unreliability factors of individual parts within components. The model was tested on 49,682 parts during 82 patrols of Polaris submarines. The data from the first 61 patrols constituted the historical period for the estimation of the parameters of the model. The data of the later 21 patrols were used as the future (test) period. The test confirms the proposition that the probability of demanding a repair part increases with the component unreliability factor.

Zacks and Zimmer [10.38] have studied the problem of estimating the factors in such multiplicative Poisson models for general reliability systems. They have derived the maximum likelihood estimators (MLE) of the severity factors and have shown that these estimators coincide with the least-squares estimators of severity factors. Bayes estimators with respect to Dirichlet prior distributions of $\rho_i = S_i/s$ $(i=1,\ldots,s)$ were derived also. The mean square error efficiency of the Bayes estimators relative to those of the MLEs was also investigated. According to Zacks and Zimmer, the estimators (10.1) proposed by Haber, Sitgreaves, and Solomon are MLEs and those given by (10.2) are best unbiased ones. Furthermore, it was shown that in certain situations, especially when the severity factors of the different patrols are not supposed to differ substantially, the Bayes estimators are more efficient than the MLEs.

In another study [10.11] Haber and Sitgreaves approached the problem of demand prediction in a different manner. They have classified the parts into classes according to the kind of part (transistor, motor, valve, and so on). In each class there are many different parts of the same kind. Their model assumes that the number of replacements (depth) of a certain part has a Poisson distribution with some mean θ. Different parts in the same class have different θ values. If we consider the variation in the θ values

as a realization of a sample from a gamma distribution $G(\alpha/\beta,\alpha)$, where α/β is the scale parameter, the marginal distribution of the number of units X demanded altogether in a given class, T patrols, will follow a negative binomial distribution, with a probability function

$$P[X=i] = \left(\frac{\alpha}{\alpha+T\beta}\right)^{\alpha} \cdot \left(\frac{T\beta}{\alpha+T\beta}\right)^{i} \frac{\Gamma(\alpha+i)}{\Gamma(i+1)\Gamma(\alpha)} \quad ,$$

$$i = 0,1,2,\ldots$$

The negative binomial distribution of the demand for parts showing movement was reported also previously in the Allowance List Test Program, by Solomon and Denicoff [10.28] and Denicoff, Fennell, and Solomon [10.5]. The parameters α and β for each class were estimated by the method of moments according to which the sample mean and sample variance in each class were equated to the expectation and variance of the negative binomial distribution. Fifty-four different classes of items were considered and in 36 of these classes the negative binomial distributions fitted the empirical distributions. The unique feature of the model is that in addition to providing positive usage estimates (of the expected values) it can be applied also to new parts that belong to a certain class, without any demand history of these new parts. As will be discussed later, these negative binomial distributions can also be used as anticipated demand distributions in Bayes procedures for adaptive control of inventory levels. At last we mention that Hadley and Whitin [10.13] have also obtained the negative binomial distribution for low demand items in an interesting model of a one-echelon multi-depot system.

10.2.2 Statistical Methods for Demand Rates Which Are Not Extremely Low. When one studies systems with demand rate that is not extremely low, a variety of forecasting methods and techniques are available. Markland [10.9] compared 6 methods of forecasting which were found appropriate for demand prediction of military helicopter spare parts. He furnished 43 references of papers in which other methods are employed. The methods that

Markland compared are the following.

1. Issue Interval. the average demand rate (per flying hour) is multiplied by the expected number of future flying hours to give an estimate of the future expected number of spare parts that will be demanded (a ratio estimate).

2. Moving Regressions. Multiple regressions were fitted by the method of least squares for the initial experience period, with the number of demands X as the dependent variable. The regressors considered were: flying hours, helicopter density over time. Various transformations of these two variables were tried. Best fitting regressions were determined according to the size of the multiple coefficient of determination R^2 . After each forecasting cycle the parameters of the regression equations were revised.

3. General Exponential Smoothing. The forecasted demand for the t-th period is

$$\hat{d}_t = AX_t/S_{t-L} + (1-A)(\hat{d}_{t-1}+T_{t-1})$$

where X_t is the actual demand during the t-th period, S_{t-L} is a multiplicative seasonal effect, L is the number of months in a cycle period, T_t is the current estimate of the linear trend effect, and A is the exponential smoothing coefficient $0 \leq A \leq 0.1$.

4. Single Exponential Smoothing. The forecast is

$$\hat{d}_{t+i} = \hat{a}_t$$

where $\hat{a}_t$ is the simple average of demand during the first t periods.

5. Double Exponential Smoothing. The forecast is

$$\hat{d}_{t+i} = \hat{a}_t + \hat{b}_t i$$

where $\hat{b}_t$ is the linear trend during the first t months.

6. Triple Exponential Smoothing. The forecast is

$$\hat{d}_{t+i} = \hat{a}_t + \hat{b}_t \cdot i + \frac{1}{2} \hat{c}_t \cdot i^2$$

where $\hat{c}_t$ is the slope of the linear trend during (θ,t) .

These procedures were compared by Markland according to the associated coefficients of variation over the forecasting horizon

$$CV_L = \frac{\delta_L}{\sum_{t=1}^{k}\left(\sum_{i=1}^{L} d_{t+i}\right)/k}$$

where $k = N - L + 1$, and

$$\delta_L = \left(\sum_{t=1}^{k} \frac{1}{k-1}\left(\sum_{i=1}^{L}\left(\hat{d}_{t+i} - d_{t+1}\right)\right)^2\right)^{1/2}$$

The numerical results obtained indicate the general superiority of the triple exponential smoothing technique over the other forecasting methods.

10.3 Adaptive Inventory Control

The literature on the stochastic control of inventory systems is voluminous. For a general review of the field one could refer to Veinott [10.29]. The various stochastic models assume the knowledge of a specific distribution function(s) of the demand random variable(s). The problem is that of statistical control

when the distribution function of the demand is unknown or not completely known. For example, suppose that the stochastic model assumes that the failure of parts in a certain system follows a Poisson process with intensity (mean number of demanded units per time unit) λ . However, the parameter λ is unknown. A common practice is to estimate the unknown parameter(s) independently of the control process and substitute the estimates of the parameters in the original control formulae. Such procedures may often prove useful. However, they require good data on the systems to which they pertain. Statistical control procedures provide methods of control that can be employed simultaneously with the collection of data. A general theoretical framework for statistical control of inventory systems is furnished by statistical decision theory. There are many papers in which a Bayesian approach has been employed. We mention here for example only a few papers: Scarf [10.26], McGlothlin [10.23], McGlothlin and Radner [10.24], Zacks [10.32], Zacks and Fennell [10.35, .36].

Scarf considered in [10.26] an inventory system in which, at the beginning of each period a decision is made concerning the stock level required. There is no backlogging and replenishment of stock is instantaneous (zero lead time). The purchase of Z units cost CZ , the holding cost is h per unit per time period and shortage cost is p per unit. It is assumed that the demand distribution belongs to a specified exponential type distribution, that is, the probability function of X is

$$P(x;w) = h(x)\exp\{-xw\}\beta(w) \quad , \quad w \in \Omega$$

The actual value of the parameter w is unknown. The optimal stock level in the case of known w is

$$S^0(w) = P^{-1}\left(\frac{p-c(1-\alpha)}{p+h} \mid w\right)$$

where α , $0 < \alpha < 1$, is a discount factor. If S_n is the actual stock level at the beginning of the nth month, the optimal ordering level is $\max\{0, S^0(w)-S_n\}$.

The paper proposes a Bayes solution for cases where w is unknown. It is shown that there exists a sequence $S_n^0(T_n)$, $n \geq 1$, where

$$T_n = \sum_{i=0}^{n} X_i \quad \text{where} \quad X_0 \equiv 0$$

is the total observed demand, such that the optimal ordering is $\max\{0, S_n^0(T_n) - S_n\}$. Properties of the $S_n^0(T_n)$ functions are derived. In particular, it is proven that the $S_n^0(T_n)$ sequence is monotone increasing. This follows from the monotone likelihood ratio property of the exponential family. An asymptotic expansion is obtained, as $n \to \infty$, for the critical functions $S_n^0(T_n)$. Under general regularity conditions it is shown that

$$S_n^0(\mu) \approx S^0(w(\mu)) + \frac{a(\mu)}{n}$$

where $\mu = E_w\{X\}$ and $w(\mu) = w$ value for which $(d/dw) \log \beta(w) = \mu$. The function $a(\mu)$ is determined explicitly. Since $w(T_n/n)$ is the maximum likelihood estimator of $w(\mu)$, the Bayes optimal policy given by $\{S_n^0(T_n)\}$ approaches the maximum likelihood estimator of $S^0(w)$, namely

$$\hat{S}^0(T_n) = p^{-1}\left(\frac{p-c(1-\alpha)}{p+h} \mid w\left(\frac{T_n}{n}\right)\right)$$

The regularity conditions on the chosen prior distribution are quite mild and we have accordingly a general large sample approximation to the Bayes solution, provided by the substitution of the maximum likelihood estimator of w in the formula of $S^0(w)$.

Zacks [10.32] considered a simple inventory system which can be adjusted at the beginning of each period (month) in order to minimize the anticipated (expected) inventory cost during that period. This adjustment allows for decreasing high stocks and increasing low stocks. It is assumed that the demand variables $X_1, X_2, \ldots$ are independent having the same Poisson distribution $P(\theta)$ where θ is unknown. The Bayesian model assumes a prior gamma distribution $G(\tau^{-1},\nu)$ for θ. It follows that after n months, the posterior distribution of θ, given the sufficient statistic T_n displayed above, is the gamma distribution $G(\tau^{-1}+n,\nu+T_n)$. The monthly inventory cost due to overstocking or to shortages is $C(S,X) = h(S-X)^+ + p(S-X)^-$, where S is the stock level at the beginning of the month (after adjustment), X is the demand during the month, h is the holding cost of a unit not demanded and p is the shortage cost per unit. Moreover, $a^+ = \max\{a,0\}$ and $a^- = -\min\{a,0\}$. Similar to the case in Scarf's model, if θ is known the optimal stock level S^0, at the beginning of each month is

$$S^0(\theta) = p^{-1}\left(\frac{p}{h+p}\ ;\ \theta\right)$$

This is the $p/(h+p)$th fractile of the Poisson distribution. When θ is unknown, it is shown by Zacks that the optimal stock level at the beginning of the $(n+1)$st month is

$$S^0_{n+1}(T_n) = \text{N.B.}^{-1}\left(\frac{p}{h+p}\ |\ \Psi_{n+1},\ \nu+T_n\right)$$

where

$$\Psi_{n+1} = \tau/(1+(n+1)\tau)$$

and $N.B.^{-1}(\alpha|p,\nu)$ denotes the α- fractile of the negative binomial distribution having a probability function

$$P[X=j|p,\nu] = \frac{\Gamma(j+\nu)}{\Gamma(\nu)\Gamma(j+1)} (1-p)^{\nu} p^{j} \quad , \quad j = 0,1,\ldots$$

This result is similar to the one mentioned earlier of Haber and Sitgreaves [10.11] concerning the determination of allowance lists according to the negative binomial marginal distribution of demand for parts that belong to the same class. We notice that the Bayes procedure yields an adaptive sequence, which strongly converges to the optimal stock level $S^0(\theta)$, whatever the value of θ is. In other papers [10.30, .33] provided a non-Bayesian adaptive procedure for determining $S^0(\theta)$, based on a sequence of uniformly most accurate tolerance limits. These procedures are based on the following idea. The optimal stock level $S^0(\theta)$ is the $p/(h+p)$th fractile of the Poisson distribution $P(\theta)$. A $(1-\alpha)$-upper confidence limit for a γ-fractile of a distribution is called a $(1-\alpha,\gamma)$ tolerance limit. As shown in [10.33] the uniformly most accurate $(1-\alpha, p/(h+p))$-upper tolerance limit after n observations is

$$S_n = P^{-1}\left(\frac{p}{h+p} \; ; \; K_{1-\alpha}(T_n)\right) \tag{10.3}$$

where $K_{1-\alpha}(T_n)$ is a uniformly most accurate $(1-\alpha)$-upper confidence limit for θ given by

$$K_{1-\alpha}(0) = -\frac{1}{n} \ell n\alpha \quad , \quad T_n = 0$$

$$K_{1-\alpha}(T_n) = \frac{1}{2n} \chi^2_{1-\alpha}[2T_n] \quad , \quad T_n \geq 1$$

We notice that $\lim_{n\to\infty} K_{1-\alpha}(T_n) = \theta$ almost surely (a.s.). This follows from the fact that $T_n/n \to \theta$ a.s. Thus,

the sequence $\{S_n\}$ given by (10.3) either converges a.s. to $S^0(\theta)$, or fluctuates between $S^0(\theta)$ and $S^0(\theta) - 1$ whenever

$$\theta = \frac{1}{2}\chi^2_{\frac{p}{h+p}}[2j+2] \quad \text{for some } j = 0,1,\ldots$$

Notice that this procedure can be applied also in cases of parts with extremely slow demand, because S_n is positive even if $T_n = 0$. Furthermore, the presently described procedure does not depend on any assumptions concerning a prior distribution of θ.

It should be remarked, however, that if α is small the tolerance limits adaptive approach will generally yield procedures that are too conservative at the beginning. That is, the corresponding stock levels S_n are too large compared to the sequence of stock levels provided by the appropriate Bayes procedures.

The Bayesian and the non-Bayesian adaptive procedures described above can be generalized also to more complicated inventory models. In [10.35, .36] Zacks and Fennell developed Bayes adaptive control procedures for two-echelon inventory systems. These procedures provide algorithms for the optimal ordering of the lower echelon stations from the upper echelons, based on the objective of minimizing the total prior expectation of the lower echelon cost. On the other hand, in order to avoid a tedious dynamic programming determination of the upper echelon's ordering levels, a Bayes prediction policy was adopted for the upper echelon.

This Bayes prediction policy determines a reordering level, which is the smallest quantity required in order that the posterior probability of satisfying the lower echelon anticipated demand will exceed a tolerance level α, $0 < \alpha < 1$. The computations become quite complicated. Algorithms and computer programs are available.

10.4 Operational Readiness

The problem of measuring and making statistical inference on operational readiness has not attracted much attention. The papers that studied this problem considered the model of a two-state Markov chain, "up" and "down." The system is in a state of readiness when it is "up." The problem is to estimate the probability of readiness as defined in the following development.

Gaver and Mazumdar [10.9] and Mazumdar [10.22] studied the problem in the context of a stationary (homogeneous) Markov process. The system starts at state U (up) and stays at this state for a random length of time X_1 then it switches to state D (down) and stays there for a random length of time Y_1. From D it switches to U and from U to D alternately. Thus, if one comes to observe the system when it is at U, he will observe a sequence of independent random variables $X_1, Y_1, X_2, Y_2, \ldots$. All the X's are exponentially distributed and the Y's are exponentially distributed. $X \sim G(\mu,1)$ and $Y \sim G(\lambda,1)$. The steady-state probability of U is defined as the operational readiness. This parameter is $P = \lambda/(\lambda+\mu)$.

The operational reliability is defined as

$$\rho = \frac{\lambda}{\lambda+\mu} \exp\{-\mu t\}$$

Gaver and Mazumdar derived the maximum likelihood (M.L.) and the uniformly minimum variance unbiased (N.M.V.U.) estimators of P and ρ, and compared the mean square errors of these estimators by Monte Carlo simulation. There is an interesting problem of sampling design, namely, when and for how long observations should be taken. Gaver and Mazumdar proposed a "patch-snapshot" design. This sampling method consists of a sequence of observations made continuously on intervals. These observations are called "patches." In addition, isolated observations are interlaced between the patches at widely dispersed instants. These observations are called "snapshots." The authors do not explain why one

should consider patches designs reinforced with snapshots if the model of stationary process prevails. There may be, however, systems which impose technical constraints so that observations cannot be taken continously over a long time interval, but only on short patches. Such a design would be necessary for control purposes, if there is a possibility that the parameters of the system may change at unknown time points.

Zacks [10.31] considered a two-state Markov chain that may change its state at the beginning of each period (day) and stay in that state during the prescribed period. The two states that are considered are U (up) and D (down). Observations on the state of the system are done regularly at the beginning of each period. The main problem under consideration is that the system may shift from one matrix of transition probabilities to another at unknown epochs. Also, the exact values of the transition probabilities are unknown. Thus, the study of Zacks provides a method of estimating the current values of the operational readiness without assuming that the Markov process is time homogeneous.

The operational readiness of a system at time period t is $\theta(t)$ and is defined as the probability that the system will be in state U during the t-th time period. Zacks' method of estimating the current operational readiness parameter is developed within a Bayesian framework which assumes at most one shift, from a matrix of transition probabilities Θ to a matrix Φ where

$$\Theta = \begin{pmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{pmatrix} \qquad \Phi = \begin{pmatrix} \phi_{00} & \phi_{01} \\ \phi_{10} & \phi_{11} \end{pmatrix}$$

The epoch of shift J from Θ to Φ is considered as a random variable having a prior discrete distribution over the set of nonnegative integers $\{0,1,\ldots\}$. Let us adopt the following notation: $\{J=0\}$ means that the shift has occurred before we have started to observe the system; $\{J=j\}$, $j = 1,\ldots,n-1$, means that the shift occurred right after the jth observation; $\{J=n\}$ is the event that the shift has not yet occurred. The

values of θ_{00} , θ_{10} , ϕ_{00} and ϕ_{10} are given prior uniform distributions on (0,1) with the assumption of prior independence. These assumptions can be obviously replaced by other, maybe more suitable, assumptions. Bayes estimators of the current transition probabilities are derived. On the basis of the derived posterior distributions Bayes estimators of the current operational readiness parameter, $\theta(t)$, are derived. Furthermore, Bayes estimators of the expected number of periods during which a system stays in a state of readiness are also given.

10.5 Replacement When A Constant Failure Rate Precedes Wearout

The planning of stock levels of spare parts depends as we have seen on the distribution of the demand for these parts. Consider a system with parts having a constant failure rate. That is, the life distributions of these parts are exponential with certain failure rates, $\lambda_1,\lambda_2,\ldots$. Suppose that the inventory system of spares for these is designed in an optimal manner relative to the specified λ-values. Systems enter generally at certain epochs to a wearout phase. This is a phase in which the failure rate is increasing. If these epochs of shift in the failure rate regime are known then the inventory system can be modified in due time to take care of the increasing demand for spare parts. Hunter and Proschan [10.15] studied such reliability problems where the wearout epochs are known. In systems that they considered the replacement of parts takes place exactly after failure or at the epoch of wearout, t, whichever comes first. Hunter and Proschan derived, by using standard renewal theory, the distributions and the expected values of the number of planned replacements, and of the number of removals due to failures or due to planned replacement. Examples of applications are given also.

An important and more realistic problem is, however, how to plan the spare parts management when the epoch of shift t is unknown. Lorden and Eisenberger [10.20] and Zacks [10.34] studied this wearout problem for unknown epochs of shift in the framework of the quickest detection theory. Lorden and Eisenberger consider a

model in which a sequence of independent random variables $X_1, X_2, \ldots$ are distributed exponentially, that is, $X_1 \sim G(\lambda_i, 1)$, where $\lambda_1 = \lambda_{m-1} = \lambda$ (known) and $\lambda_m = \lambda_{m+2} = \lambda_{n+2} = \ldots = \lambda(1+\theta)$ where θ, $0 < \theta < \infty$, is also known. The problem of deciding whether the system has entered a wearout phase can be considered as a problem of choosing a stopping variable N. This is a random variable that assumes positive integer values and for each $n = 1,2,\ldots$ the event $\{N \geq n\}$ is measurable in the sigma-field generated by the first n observations. Whenever N is realized, the part has to be replaced by a new one or some other rectifying action should take place. Lorden and Eisenberger develop the following criterion for a "good" detection procedure. Consider the conditional expectation, under m of the excessive number of observations, that is, $E_m\{N-(m-1) \mid X_1, \ldots, X_{m-1}\}$, where $\{X_1, \ldots, X_{m-1}\} \in \{N \geq m\}$. Define for each $m \geq 1$

$$c_m = \sup \left\{ E_m\{N-(m-1) \mid X_1 = x_1, \ldots, X_{m-1} = x_{m-1}\} ; \right.$$
$$\left. \{x_1, \ldots, x_{m-1}\} \in \{N \geq m\} \right\}$$

Furthermore, define for a given θ, and a stopping variable N,

$$\bar{E}_\theta\{N\} = \sup_{m \geq 1} c_m$$

The objective is to choose a stopping variable N having a finite expectation $E_\theta\{N\} < \infty$, for which $\bar{E}_\theta\{N\}$ is minimized over a specified range $[\theta_1, \theta_2]$. Without a constraint the procedure may yield too many "false alarms." Thus, Lorden and Eisenberger impose the constraint that $E\{N\} \geq \gamma > 1$, for some specified γ. $E_0\{N\}$ is the expected stopping time when there is no shift ($\theta=0$). The performance of a stopping variable proposed by Page [10.25] in a control charts context was also investigated. Page's stopping variable can be

defined in the following manner.

N = least positive integer n such that

$$T_n \geq \log \gamma \text{ , where } T_0 = 0 \text{ and} \qquad (10.4)$$

$$T_n = \max\{0 \text{ , } T_{n-1} + \log (1-\theta) - \theta X_n\} \text{ .}$$

It is shown that Page's procedure is asymptotically, as $\gamma \to \infty$, optimal in the above sense. Approximations for the expected values of N which is defined in (10.4) are provided and compared to Monte Carlo estimates. A modification of the procedure is proposed for the case when, in addition to the unknown epoch of shift m , the scale parameter λ is also unknown. It is shown that if $S_n = X_1 + X_2 + \cdots + X_n$, then

$$\log \frac{S_2}{S_1} \text{ , } 2 \log \frac{S_3}{S_2} \text{ , } 3 \log \frac{S_4}{S_3} \text{ , } \ldots$$

is a sequence of independent exponentially distributed random variables, independently of the unknown λ . The Page procedure is then applied on the sequence of random variables $W_1, W_2, \ldots$ where $W_n = n \log (S_{n+1}/S_n)$. After the epoch of change the distribution of W_n is like that of Weibull random variables with a shape parameter $\alpha = 1/(1+\theta)$ and arbitrary scale parameter. The Weibull distributions have increasing or decreasing failure rate functions according to the value of α . Thus, the above transformations reduce the problem from that of detecting the epoch of shift in the scale parameter λ to a problem of detecting the epoch of shift from an exponential distribution to a Weibull distribution.

Zacks [10.34] studied the problem of detecting an epoch of shift from an exponential to a Weibull distribution in a Bayes adaptive framework. It is assumed that the failure rate function is

$$h(t;\tau) = \lambda \text{ , if } t \leq \tau$$

$$= \lambda + \lambda\alpha(t-\tau)^{\alpha-1} \quad , \text{ if } \quad t > \tau$$

where τ is the epoch of shift and α is the shape parameter of the Weibull distribution. It is assumed that τ is a random variable having a prior continuous distribution, with a density function $\xi(\tau)$. The observed random variables $X_1, X_2, \ldots, X_n, \ldots$ are truncated exponential random variables. The values of X are the minimum of the failure time or planned replacement time of a part. After observing $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ the posterior probability of $\{\tau > t_n\}$ given $\underset{\sim}{X}_n$, where $t_n = X_1 + X_2 + \cdots + X_n$, is computed. As long as this posterior probability Π_n is greater than a pre-assigned level Π^*, the replacement scheme continues as though the shift has not yet occurred. The switching to the replacement policy appropriate for the Weibull distribution takes place as soon as $\Pi_n < \Pi^*$. A computing algorithm for the posterior probabilities Π_n, under a prior exponential distribution of τ, is given. The operating characteristics of the procedure are studied by Monte Carlo simulation. There are still several open questions for further research.

(1) The relationship between the critical level Π^* and the expected number of false alarms, on the one hand, and the expected delay in detection on the other hand

(2) The sensitivity of the Bayes procedure with regard to the choice of the prior distribution of τ

(3) The extension of the results to cases where the scale and shape parameters are also unknown

(4) The comparison of the (relative) efficiency of the present Bayes procedure to that of Page

Harris, Marchal, Singpurwalla and Zacks [10.14] and

Goldschen and Singpurwalla [10.10] presented several other statistical methods of testing for a change in the failure rate function of a system. Some of these procedures proved by Monte Carlo estimates to be quite effective.

10.6 Surveillance

Under surveillance problems we include a class of inspection problems of reliability or production systems that deteriorate or fail in time. The objective is generally to keep the system in operational or standard condition. In other words, it is desirable to replace units or components of the system that fail immediately after their failure epochs. However, generally the inspection of the system is a costly operation and in many (situations) cases the systems are quite reliable in the sense that it may take a long (but random) time until the failure of its units. A continuous monitoring is in many cases not justified. The surveillance problem can be described in general terms as the problem of determining, or planning, the schedule of inspection epochs so that, the expected cost due to inspection and due to failure of components will be minimized. Two major optimality objectives are discussed in the literature: minimizing the total discounted expected cost for the entire future of the system, or minimizing the long-run average cost per unit time.

There are many papers in the literature that study surveillance problems or related maintenance problems under stochastic failure regimes. We mention here only a few which present some of the main approaches. One of the early papers is that of Barlow and Hunter [10.2]. For the general theory see Chapter 4 in the book of Barlow and Proschan [10.3]. Barlow and Hunter provided a solution to the problem for cases of known failure time distributions. Ehrenfeld [10.7] considered the problem from the point of view of designing binomial experiments when N units operate in parallel and have the same exponential failure time distribution. The available information at the inspection epoch is how many units among the N failed during the last period between inspections. Since the assumed model is exponential the replacement strategy is equivalent to that of replacing all the units with new ones at every

inspection epoch. The cost function typically consists of two components: the cost of inspection, which depends on the number of units or components in the system, and the cost of failure per time unit. Variations of the cost structure and of the replacement strategy in cases of failure times that are not exponential has produced many studies. We cite here in particular those of Kamins [10.16], Kander [10.17] and Kander and Naor [10.18]. Zacks and Fenske [10.37] have provided a sequential adaptive procedure of determining the epochs of inspection when the failure time distributions are general and the units are not replaced before their failure time. The objective in that paper is to minimize the average expected cost per inspection interval. After m inspections the determination of the $(m+1)$st inspection epoch is a function of the first m inspection times and of the random vector $\underset{\sim}{N}^{(m)} = (N_0^{(m)}, N_1^{(m)}, \ldots, N_{m-1}^{(m)})$ where $N_j^{(m)}$ $(j=0,\ldots,m-1)$ is the number of components that were replaced at the jth inspection epoch and are still in operation at the current inspection. In particular, the Weibull family of distributions with increasing or decreasing failure rate functions is considered.

Statistical surveillance problems arise when the distributions of failure times are either unknown or not completely known. The problem of determining optimal inspection epochs when the distributions are unknown are much more difficult. Consider for example even the relatively simple model of exponential failure time. When the mean time between failures θ is known the solution is simple and the time period between inspections remains the same as long as θ is unchanged. On the other hand, when θ is unknown the determination of the optimal inspection epochs becomes a difficult matter. One can obviously obtain Bayes determination of the inspection epochs over a finite planning horizon for a fixed number of inspections n by the method of dynamic programming, assuming a prior distribution function for the unknown parameter θ (see Kander and Rabinovitch [10.19]). The numerical procedure involved may be quite laborious. Another possibility is to consider adaptive estimation of the unknown parameter θ,

by an appropriate Bayesian or non-Bayesian procedure, which is combined with the sequential determination of the inspection epochs. Fenske [10.8] compared numerically the characteristics of such adaptive procedures based on Bayes estimators and an average maximum likelihood estimator. The two procedures proved to yield effective results and to converge quite fast to the optimal inspection interval for a given θ. However, the Bayes procedure requires a substantial amount of computer time while the average maximum likelihood method can be easily and quickly executed. The research in this type of adaptive statistical surveillance problem has only begun and it seems to be an important and promising field of research. We conclude this section with a comment that a minimax solution to the surveillance problem when the distributions are unknown is easy to obtain (see Derman [10.6]). However, the main criticism against the minimax approach is that it is not adaptive. The whole schedule can be determined ahead. Accordingly, if the number of allowed inspections and the planning horizon are large, a consistent adaptive procedure can reduce the average surveillance cost per time unit since it converges to the optimal solution for the case of known distribution. The adaptive procedures are based, however, on more specific models concerning the failure time distributions. Such models are generally justified.

References

[10.1] Astrachan, M., and A. S. Cahn (eds.) (1963). Proceedings of Rand's Demand Prediction Conference January 25-26, 1962. RM-3358-PR, The Rand Corporation.

[10.2] Barlow, R. E., and L. C. Hunter (1959). Mathematical models for system reliability. Engineering Report No. EDL-E35, Sylvania Electric Products, Mountain View, California.

[10.3] Barlow, R. E., and F. Proschan (1967). Mathematical Theory of Reliability. Wiley.

[10.4] Denicoff, M., J. Fennell, S. E. Haber, W. H. Marlow, F. W. Segal, and H. Solomon (1964). The Polaris military essentiality system. Naval Res.

Logist. Quart. 11 235-257.

[10.5] Denicoff, M., J. Fennell, and H. Solomon (1957). Requirements determination. Serial 59/57, Logistics Research Project, The George Washington University.

[10.6] Derman, C. (1961). On minimax surveillance schedules. Naval Res. Logist. Quart. 8 415-419.

[10.7] Ehrenfeld, S. (1962). Some experimental design problems in attribute life testing. J. Amer. Statist. Assoc. 57 668-679.

[10.8] Fenske, W. J. (1972). Optimal inspection epochs for reliability systems. PhD dissertation, Case Western Reserve University.

[10.9] Gaver, D. P., and M. Mazumdar (1967). Statistical estimation in a problem of system reliability. Naval Res. Logist. Quart. 14 473-488.

[10.10] Goldschen, D. Y., and N. D. Singpurwalla. (1973). On the prediction (forecasting) of failure rates. Serial TM-62128, Program in Logistics, The George Washington University.

[10.11] Haber, S. E., and R. Sitgreaves (1970). A methodology for estimating expected usage of repair parts with application to parts with no usage history. Naval Res. Logist. Quart. 17 535-546.

[10.12] Haber, S. E., R. Sitgreaves, and H. Solomon (1969). A demand prediction technique for items in military systems. Naval Res. Logist. Quart. 16 297-308.

[10.13] Hadley, G., and T. M. Whitin (1961). A model for procurement, allocation redistribution for low demand items. Naval Res. Logist. Quart. 8 395-414.

[10.14] Harris, C. M., W. G. Marchal, N. D. Singpurwalla, and S. Zacks (1973). Failure rate prediction and wearout detection. Serial TM-63646, Program in Logistics, The George Washington University.

[10.15] Hunter, L., and F. Proschan (1961). Replacement when constant failure rate precedes wearout. Naval Res. Logist. Quart. 8 127-136.

[10.16] Kamins, M. (1960). Determining checkout intervals for systems subject to random failures. RM-2578, The Rand Corporation.

[10.17] Kander, Z. (1971). Inspection policies of deteriorating equipment characterized by N quality levels. Operations Research Monograph No. 93, Technion-Israel Institite of Technology.

[10.18] Kander, Z., and P. Naor (1970). Optimization of inspection policies by dynamic programming. Operations Research Monograph No. 61, Technion-Israel Institute of Technology.

[10.19] Kander, Z., and A. Rabinovitch (1972). Maintenance policies when failure distributions of equipment is only partially known. Operations Research Monograph No. 92, Technion-Israel Institute of Technology.

[10.20] Lorden, G., and I. Eisenberger (1973). Detection of failure rate increases. Technometrics. 15 167-175.

[10.21] Markland, R. E. (1970). A comparative study of demand forecasting techniques for military helicopter spare parts. Naval Res. Logist. Quart. 17 103-119.

[10.22] Mazumdar, M. (1969). Uniformly minimum variance unbiased estimates of operational readiness and reliability in a two-state system. Naval Res. Logist. Quart. 16 199-206.

[10.23] McGlothlin, W. H. (1963). Development of Bayesian parameters for spare-parts demand prediction. RM-3699 PR, The Rand Corporation.

[10.24] McGlothlin, W. H., and R. Radner (1960). The use of Bayesian techniques for predicting spare parts

demand. RM-2536, Rand Corporation.

[10.25] Page, E. S. (1954). Continuous inspection schemes. Biometrika 41 100-115.

[10.26] Scarf, H. (1959). Bayes solutions of the statistical inventory problem. Ann. Math. Statist. 30 490-508.

[10.27] Solomon. H. (1962). A summary of the Logistics Research Project's experience with problems of demand prediction. in [10.1].

[10.28] Solomon, H., and M. Denicoff (1960). Simulations of alternative allowance list policies. Naval Res. Logist. Quart. 7 137-149.

[10.29] Veinott, A. F., Jr. (1966). The status of mathematical inventory theory. Management Sci. 12 745-777.

[10.30] Zacks, S. (1968). Uniformly most accurate upper tolerance limits in the Poisson case. Technical Report #2, NSF Project GP-9007, Department of Mathematics and Statistics, University of New Mexico.

[10.31] Zacks, S. (1969a). Bayes adaptive estimation of current operational readiness parameters. Serial TM-61018, Program in Logistics, The George Washington University.

[10.32] Zacks, S. (1969b). Bayes sequential design of stock levels. Naval Res. Logist. Quart. 16 143-155.

[10.33] Zacks, S. (1970). Uniformly most accurate upper tolerance limits for monotone likelihood ratio families of discrete distributions. J. Amer. Statist. Assoc. 65 307-316.

[10.34] Zacks. S. (1972). On detecting the epoch at which the failure rate of a reliability system starts to increase. Serial TM-80320, Program in Logistics, The George Washington University.

[10.35] Zacks, S., and J. Fennell (1972). Bayes adaptive control of two-echelon inventory systems, I: Development for a special case of one-station lower echelon and Monte Carlo evaluation. Naval Res. Logist. Quart. 19 15-28.

[10.36] Zacks, S., and J. Fennell (1974). Bayes adaptive control of two-echelon inventory systems, II: The multi-station case. Naval Res. Logist. Quart. 21 575-593.

[10.37] Zacks, S., and W. J. Fenske (1973). Sequential determination of inspection epochs for reliability systems with general life time distributions. Naval Res. Logist. Quart. 20 377-386.

[10.38] Zacks, S., and W. J. Zimmer (1972). Estimators of severity factors in a multiplicative Poisson model. J. Amer. Statist. Assoc. 67 192-195.

Chapter 11

A SURVEY OF INVENTORY THEORY AND PRACTICE*

Donald Gross
The George Washington University
and
David A. Schrady
Naval Postgraduate School

11.1 Introduction

Inventory theory is commonly said to have begun with the development of the economic lot size formula of F. Harris in 1915 [see Harris (1915)]. The same formula was apparently independently developed by R. H. Wilson and W. A. Mueller in 1926 and often goes by the name of "economic order quantity" (EOQ), or Wilson's lot size formula [see Wilson and Mueller (1926-7)]. This formula and variations of it are still in common usage today. Since that time and especially post 1950, a voluminous body of theoretical inventory literature has evolved. Even so, the EOQ formula still enjoys great popularity. What does this, then, imply about the implementability and general usefulness of the massive theoretical developments of the 1950s and 1960s? Is there a gap between theory and practice and if so, how large is it? We will attempt to provide some insight concerning this question as we survey both the theoretical developments and the policies in common usage today.

One can conveniently display the "levels of effort" in inventory control by a Venn diagram as shown in Figure 11.1. The set T represents the theoretical development in inventory control policies, while the set R represents the real world of inventory policies. The policies used in the real world must be implementable and computationally feasible. Many computationally feasible policies in use are not derived from any

*The first author was supported in the preparation of this chapter by the Office of Naval Research under Contract N00014-67-A-0214-0001 with the George Washington University.

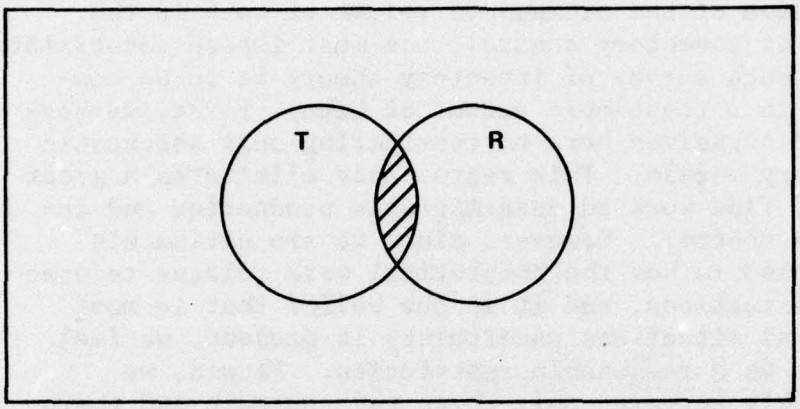


Figure 11.1 - A Venn diagram.

theory (for example, keep thirty days of supply on hand for every item). The intersection $T \cap R$ (cross-hatched area) represents those policies used in practice that are computationally feasible and actually emanated from the theory. Thus, the gap between theory and practice could be measured by

$$\text{Gap} = \frac{1 - (T \cap R)/R}{(T \cap R)/R}$$

so that if $(T \cap R)/R = 1$, Gap = 0, while if $(T \cap R)/R = 0$, Gap = ∞. Hence we desire a feel for the relative magnitude of $(T \cap R)/R$, and hope to gain such a feel as we survey the theory and practice of inventory control.

The organization of the chapter is as follows. Section 11.2 gives the theoretical survey while Section 11.3 investigates the policies in common usage, with a detailed look at the military. Finally, in Section 11.4 we return to the discussion of the Gap.

11.2 Theoretical Survey

Because of the tremendous volume of work in the theory of inventory control, one must impose constraints if any such survey of inventory theory is to be completed in a reasonable amount of time. First, we have limited ourselves here to considering only stochastic inventory models. This regrettably eliminates a great deal of fine work on deterministic production and inventory control. However, since we are ultimately interested in how the theoretical work relates to practical situations, and it is our belief that in most practical situations uncertainty is present, we feel this to be a reasonable restriction. Second, we thoroughly surveyed only three key journals and therefore the bulk of our survey is comprised of articles published therein, although articles from other journals are included when the bibliography of key references and our own knowledge led us elsewhere. The three key journals surveyed are Operations Research, Management Science and Naval Research Logistics Quarterly, which we believe to be representative of theoretical work in the field of inventory logistics. Third, certain areas of special concentration were also omitted; for example, we omit the interesting work on perishable inventory models.[1] Again, we find these types of omissions regrettable but believe them to be necessary for the sake of expediency. Finally, there are undoubtedly inadvertent omissions even within our own constraint set, so that the references surveyed should be thought of as a sample (hopefully large and representative) of stochastic inventory theory.

There are numerous ways in which one could attempt to categorize inventory models. For our purposes here, we concentrate on the following three: (1) type of policy; (2) solution approach, and (3) results obtained.

---

[1]For typical references in this area, see S. Nahmias (1974), Inventory depletion management when the field life is random, Management Sci. 20 1276-1283; W. P. Pierskalla and C. Roach (1972), Optimal issuing policies for perishable inventory, Management Sci. 18 603-614; and C. C. Pegels and A. E. Jelmert (1970), An evaluation of blood bank inventory policies: a Markov chain application, Operations Res. 18 1087-1098.

There are two basic types of inventory control policies, namely, periodic review (P) or continuous review (C). In the former, inventory is reviewed on a periodic basis (the period generally being deterministic) and the order quantity, which depends upon the inventory position at review, is a random variable. Work on P policies can be further subdivided into single period (static) cases (P_1), or (dynamic) finite horizon (P_n), or infinite horizon (P_∞) cases. Generally, the decision variables to be determined are the inventory position level, which indicates if an order should be placed, the amount to order (or equivalently the inventory position desired after ordering) if an order is to be placed, and sometimes the length of the review period itself. These periodic situations lead quite naturally to (s,S) type rules, which indicate an order is to be placed to bring inventory position up to S if upon review the inventory position is below s. Much of the theoretical work on P policies involves the study of the characteristics of such rules and the necessary conditions (on costs, demand densities, and so on) that yield (s,S) policies as optimal. The criteria of optimality are generally the total costs (profits) for P_1 and P_n cases and either the sum of the discounted cost (profit) stream or the average cost (profit) rate for P_∞ situations. In some P_n cases, discounting is also considered.

For C policy situations, inventory position is continuously monitored so that the decision variables are the level of inventory position that triggers an order and the amount to order. Here the order quantity is deterministic but the time between placing orders is a random variable. These can be looked at as continuous review (s,S) policies, where s denotes the trigger point and the order quantity is given by S-s; however, it is common to see them referred to as (r,Q) policies, where r and Q denote the trigger point and order quantity, respectively (r=s, Q=S-s). The criterion of optimality for C models is generally the average cost (profit) rate over an infinite horizon. Very little work has been done on finite horizon

C policies and for the purposes of our study here, C policies will imply an infinite planning horizon.

Essentially for all P_∞ and C situations, conditions on costs, demand, etc., must be assumed stationary (time-independent) by the very nature of considering an infinite planning horizon, although Bather (1966) does attempt to model a nonstationary C situation by using a Wiener process to describe demand.

Generally, three cost functions are considered to make up the cost (profit) optimization criterion, namely, a cost of procurement, a cost of holding inventory, and a shortage cost. The most common assumptions for these cost functions are linear (with an additional fixed portion for ordering cost also quite frequently considered) although a sizable number of papers have dealt with generalizations of these cost functions (concave, convex, and so on). Shortages are assumed to be back-ordered in about two-thirds of the papers while the other one-third considers lost sales.

Not all studies consider optimizing pure costs or profit functions of the decision variables. Lately, more emphasis has been put on constrained optimization, for example, minimizing ordering plus holding costs subject to service constraints (service having to do with back-order or lost sales amounts), or minimizing shortages subject to budget constraints. In a few papers, such things as the structure of the stochastic process or the stability of certain policies from a control point of view were considered. However, in the majority of the studies, one does find as the criterion of concern the optimization of costs (or profits).

With respect to assumptions concerning demand, the most common single assumption has been Poisson, although numerous papers work with general demand distributions. One finds that general demand distributions (no specific assumptions) are more common with P policies while for C pclicies the Poisson or compound Poisson is almost always used. A number of the early papers on P_∞ models also considered exponential density functions for demand.

So far we have neglected any mention of the important parameter lead time. In most P situations, lead time is treated as deterministic and quite often as zero

(instantaneous replenishment) while for the C cases, stochastic lead times are somewhat easier to incorporate, and lead times are treated as such in many of the articles.

Turning attention now to the second categorization factor (solutions approach), four major approaches appear in the theoretical literature. The first we refer to as the direct approach, where expressions in terms of the decision variables are derived from basic probabilistic considerations and, if optimization takes place, differential calculus, methods of finite differences, or numerical search procedures are employed. The second major approach is the recursive analysis of dynamic programming, the third a renewal theoretic approach, and the fourth a Markov process analysis.

Subjective judgment is sometimes required in categorizing the approach an author actually used. For example, in some approximate C models, the average costs over a cycle can be written down directly and then divided by the average cycle time to yield an average cost rate. While we consider this a direct approach, it does require the renewal-reward theorem of renewal theory to give it legitimacy. We do not go into any details of the different approaches here but refer the reader to any of the excellent survey articles listed in the references, particularly Scarf (1963), Veinott (1966) and Iglehart (1967). Although these are somewhat out of date, no new methodological developments have been achieved since, except possibly for the work of Veinott on lattice programming, which to the best of our knowledge, has not been published as yet.

Our third and final major categorization involves the types of results. Results are generally of four types: (1) concise analytical formulas; (2) numerical results such as those from search algorithms or dynamic programming; (3) general results involving integral equations, transforms, and so on; and (4) policy characterizations such as necessary conditions for optimality of (s,S) policies. Again, some subjectivity is used since a single paper can have several types of results. For example, a paper may basically present general results, but for some specialized cases (say exponential demand) a formula or two may also be given. For the "statistical analyses" to follow, we subjectively

weighted cases such as these.

Another factor explicitly considered in our analysis of theoretical work is whether any attention is given to demand estimation. Most of the work in inventory can be considered as risk modeling, that is, the demand probability distribution is assumed to be known. Some papers, however, do consider the problem of decision-making under uncertainty, usually treating this by (1) a decoupled forecasting procedure such as maximum likelihood or exponential smoothing plugged into the risk inventory model, (2) a Bayes forecasting procedure imbedded into the model itself, or (3) a min-max approach. The decoupled approach is the most common, followed next by the Bayes approach. Only a very few early papers considered the min-max approach.

Finally, of interest also is whether the authors treat single-item single-echelon problems, multi-item problems or multi-echelon problems.

Our intent in this theoretical survey is not to present an annotated bibliography or paragraph description of each reference—time and space prohibit this. We do desire, however, to gain an overview of the total work done and thus are interested in summary information. Specifically, our goals for this section of the paper are the following.

1. To ascertain any time trends in
   a. Total research effort
   b. Type of research effort
2. To investigate the relationship between
   a. Policy type and type of approach
   b. Policy type and types of results
   c. Type of approach and types of results
3. To provide information for comparison with Section 11.3 to gain insight as to the relative magnitude of $T \cap R$.

The statistical analyses to follow are based on the 159 research articles and monographs listed in the bibliography, with the exception of Harris (1915) and Wilson and Mueller (1926-7), making a total sample size

of 157. Also listed in the bibliography, under separate headings, are survey articles and textbooks that are not included in the sample used for the statistical analyses. We first investigate time trends in inventory research. Figures 11.2 through 11.6 are graphs of number of articles, policies treated, approaches used, and results obtained versus year.

In viewing Figure 11.2, in which is plotted the number of articles in the sample versus the year in which they appeared, a sizable effort begins to appear in the late 1950s [stimulated by the classic paper of Arrow, Harris and Marschak (1951)] and continues through the Sixties and Seventies. This may be due in part to sample bias favoring the current years, but even taking this into consideration, effort on inventory research does not seem to be on the wane as many have claimed. The emphasis is changing, however, and this will show up in the subsequent analysis. The year 1958 stands out due to the publication of a series of articles (each treated separately in the sample) in the classic book, Studies in the Mathematical Theory of Inventory and Production, edited by Arrow, Karlin and

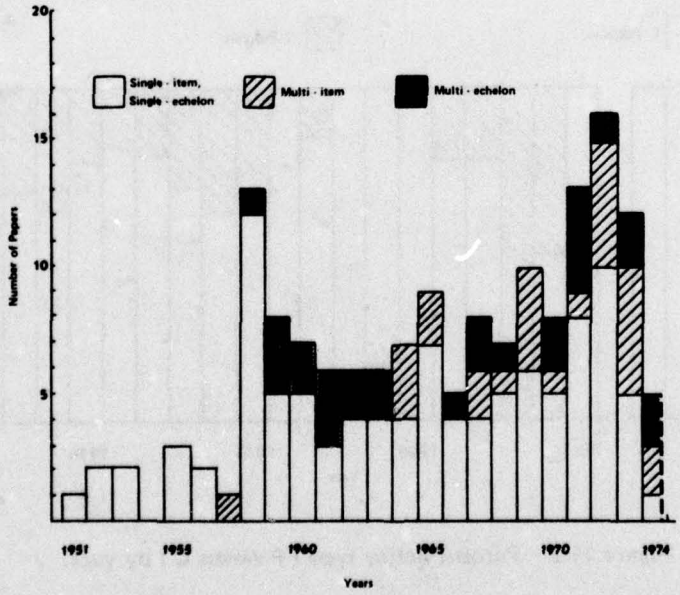


Figure 11.2 - Number of papers versus year.

Scarf, Stanford University Press, 1958. This book also served as a stimulus to further research. In Figure 11.2, multi-item and multi-echelon research is broken out separately. Effort in these areas began in 1957-1958, and continues, although it represents only 35% of the total effort, with multi-echelon representing 20% and multi-product representing 15%. The late 1960s and 1970s show an increase in multi-product effort. These problems are difficult but, of course, more realistic than single-item single-echelon problems.

Figure 11.3 shows, by year, the precentage effort in P versus C policies. The P policies represent about 77% of the total research effort to date. The early stochastic inventory work was entirely devoted to P policies, with C policies coming on the scene in the latter half of the 1950s. The early Sixties show a return to P policies, followed by a small but consistent effort on C policies from the latter Sixties to the present, with some evidence of an increasing trend in effort devoted to C policies. This could be due to one of two reasons: (1) the large historical effort in P policies leaves more research

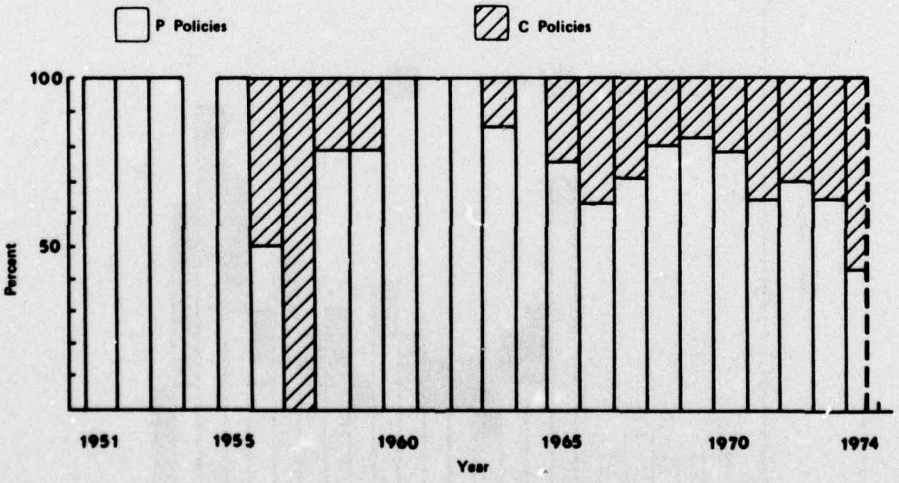


Figure 11.3 - Percent policy type (P versus C) by year.

potential in C policies, or (2) the increase in computer capacity and decrease in computing cost make transaction reporting systems a realistic phenomenon and the research is following the practice.

In Figure 11.4, a breakdown of the P policy effort with respect to P_1, P_n and P_∞ is given. The total P policy effort over all years breaks down to approximately 21% for P_1 policies, 44% for P_n policies, and 35% for P_∞ policies. The P_1 policies had their greatest effort in the early days (these were relatively "easy" to treat), with P_∞ policies enjoying their greatest popularity in the late Fifties and early to middle Sixties. The P_n policies enjoy a rather constant popularity but, as we mention below, a change in research attitude appears in later years.

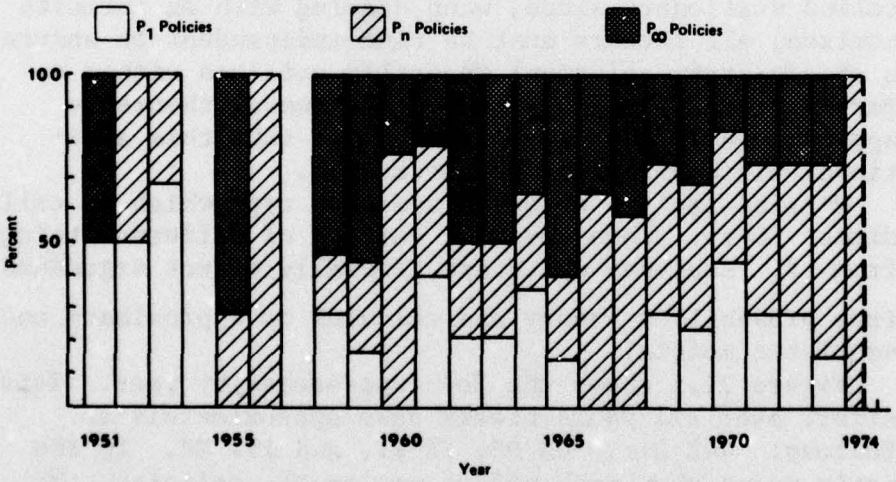


Figure 11.4 - Percent P policy type by year.

We now turn our focus on the type of approach. Professor Iglehart in an unpublished report[2] states, "... two principal analytic techniques used to study inventory models: dynamic programming and the stationary analysis of a fixed class of ordering rules." These two approaches are also presented in Iglehart's published survey [see, under survey articles, Iglehart (1967)]. The principal philosophical difference between these two methods is as follows. The recursive approach of dynamic programming (we label this DP) makes no assumptions as to the form of the policies and attempts to characterize the conditions necessary to ensure that simple policies are optimal and/or provide a means of calculating optimal values for the two decision variables, when and how much to order. The stationary analysis approach assumes a simple form for a policy [say (s,S)] and concentrates then on how to find the optimal values s and S over an infinite planning horizon. This stationary analysis (it is called stationary since, when dealing with an infinite horizon, all factors must be time-independent to ensure a steady-state solution) generally utilizes either a Markov process analysis (MP) or a renewal theoretic approach (RT). We have attempted to make this additional differentiation in our sample.

We also include a fourth approach type which we call direct (Dir). This covers a variety of different things from P_1 analyses which are generally direct arguments from probability theory and calculus to approximate and heuristic models.

Figure 11.5 shows the four approaches by year. Total effort over all years breaks down approximately as follows: 44% Dir, 30% DP, 7% RT, and 19% MP. In the early years when much effort was on P_1 policies, the Dir approach often appears. From the mid-1950s on, MP seems to be rather consistent and is used both for P_∞

[2] Iglehart, D. L. (1969). Recent developments in stochastic inventory theory. Technical Report No. 9, Department of Operations Research, Stanford University, p. 1.

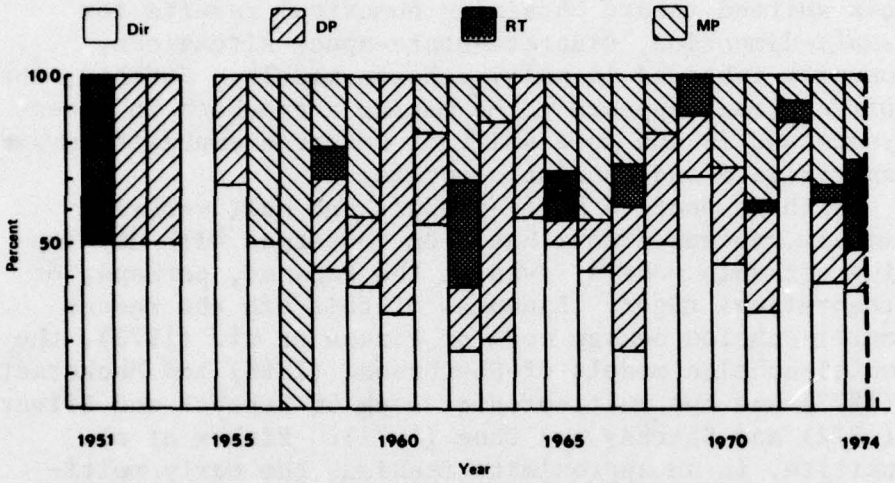


Figure 11.5 - Type of approach by year.

models, where the state probabilities of inventory level at the beginning (or end) of each period are found via a discrete paramater Markov-chain analysis, and for C policies, where the inventory level at any time t (in steady state) is found by an analysis quite similar to that used in queuing theory (continuous parameter Markov chain). In fact, most C policies can be analyzed by changing them into an equivalent queuing model, with the queuing system corresponding to the order processing portion of the inventory system. For both P_∞ and C policies, once the inventory level probabilities are determined, a cost structure is superimposed and optimization then takes place.

The RT effort is the smallest and has been used for the most part for P_∞ policies, where the expected cost over a cycle (defined as the time between successive inventory levels of a specific quantity, for example, S) and expected cycle time are developed with the aid of the renewal function and key renewal theorems. Then the renewal-reward theorem is invoked to allow division of cost per cycle by cycle time to

obtain average cost rate. Dynamic programming was very popular in the Fifties and Sixties when a great effort was placed on characterizing conditions necessary for optimality of (s,S) policies. The recent DP effort has shifted toward obtaining numerical results for small-dimension, discrete-state-space situations, usually imbedded in multi-echelon models. Further, the Dir approach appears to be making a comeback in recent years and is due to a shift of research emphasis toward approximate and heuristic models.

With respect to multi-product and multi-echelon models, recent effort has been concerned with developing workable models, even at the expense, perhaps, of theoretical rigor. Examples of this are the recent multi-echelon design work of Pinkus et al. (1973), the multi-echelon models of Sherbrooke (1968) and Muckstadt (1973) and the multi-product work of Schaack and Silver (1972) and Schrady and Choe (1971). Pinkus et al. utilize, in an approximate fashion, the early multi-echelon work of Clark (1960) and Clark and Scarf (1962), which results in a DP calculation of P_n policies. Schaack and Silver consider a multi-product C policy with a joint ordering provision for items supplied by a vendor common to the one supplying an item that has reached its trigger point. Schrady and Choe consider a nonlinear programming model of a multi-product C policy so as to minimize shortages subject to budget and ordering frequency constraints.

The above discussion quite naturally leads us into consideration of the types of results emanating from the research. We have characterized types of results as analytical (closed form solutions in the form of relatively simple formulas), numerical (iterative or algorithmic procedures), general (transforms, integral equations, and so on) and characterization (conditions needed for optimality of simple policies such as s,S). The abbreviations we use are a , n , g and ch , respectively. Figure 11.6 shows results by type versus year. The thick black line divides a+n from g+ch and can be looked at as a kind of dividing line between results potentially useful in practice (a+n) and results mainly of theoretical interest (g+ch) . However, it can be argued that some of the a and n results are for unrealistic policies while the knowledge

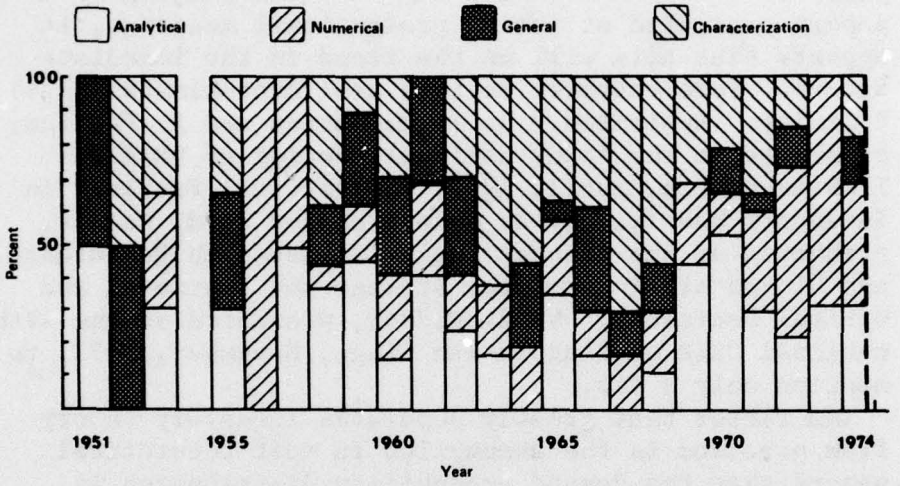


Figure 11.6 - Type of results by year.

that (s,S) is an optimal policy (ch type result) is practically useful since the practitioner must find only two values and perhaps could employ simulation to do so. Thus, once again we remark that any attempt at analyses of the kind we are doing in this section is intended to provide only a general impression and is the reason we placed "statistical analyses" in quotes when we first introduced the term.

The total effort over all years breaks down as follows: 30% a, 28% n, 16% g, and 26% ch. Thus, over half of the results are comprised of a and n , so we might consider about 60% of the research effort as potentially of direct use in practice. In reality, we feel the results actually used are considerably smaller than this since many of the computationally feasible policies ignore factors that must be considered in real life situations.

The percent a+n results are lowest during the 1960s when the major emphasis was on characterization of conditions for optimality, much of this work being the result of doctoral dissertations. There seems to be a swing back now (beginning in 1968 and apparently

continuing) to effort that produces computationally feasible policies. As previously pointed out, much of the recent work has been on approximate and heuristic policies that yield actual numbers, and judging by the papers presented at recent professional meetings, it appears that this will be the trend in the immediate future. Some examples of this are "Approximate (s,S) Policies: Revisited," by T. E. Morton and D. Pentico, presented at the joint national meeting of ORSA and TIMS in Boston, April, 1974; "Statistical Problems in Inventory Models," by A. McCormick and H. M. Wagner, presented at the present Logistics Research Conference, and "Inventory Information Storage and Retrieval and Optimal Decisions," by E. Naddor, presented at the 44th national ORSA meeting in San Diego, November, 1973, to mention only a few.

One factor that greatly separates inventory theory from practice is the assumption in most theoretical papers that the demand probability distribution is known (that is, decision-making under risk rather than under uncertainty). Only 11% of the papers in the sample treat the problem of estimating demand. Generally, the approach is to use some type of forecasting scheme to estimate demand and then, treating it as if it were known, utilize a risk inventory model--a sort of decoupled approach. The other approaches are a Bayesian forecasting procedure imbedded in the inventory model, and a min-max approach. The most common approach is the decoupled one (44% of those papers treating uncertainty), while the other two approaches split the remainder at 28% each. Min-max is generally found in the earlier work and associated with P policies. Some Bayes work appears in certain of the multi-echelon research, for example Sherbrooke (1968) and Zacks and Fennell (1972).

We conclude our statistical analyses by considering the interaction among policy type, approach type and types of results obtained. In Figure 11.7 we present three matrices: (a) policy type versus approach, (b) policy type versus results, and (c) approach versus results. In each cell there are two numbers separated by a diagonal. The lower left number is associated with a row while the upper right number is associated with a column. The numbers represent percentages. For

(a) Policy versus approach

Policy \ Approach	Dir	DP	MP	RT	
P_1	33 / 100				16
P_n	25 / 33	65 / 66		7 / 1	34
P_∞	22 / 32	32 / 36	24 / 16	60 / 16	27
C	20 / 30	3 / 4	76 / 56	33 / 10	23
	44	30	19	7	

(b) Policy versus results

Policy \ Results	a	n	g	ch	
P_1	19 / 39	13 / 21	9 / 11	14 / 29	16
P_n	11 / 10	35 / 28	32 / 16	47 / 46	34
P_∞	33 / 34	14 / 12	34 / 20	32 / 34	27
C	37 / 41	38 / 35	25 / 16	7 / 8	23
	30	28	16	26	

(c) Approach versus results

Approach \ Results	a	n	g	ch	
Dir	60 / 40	44 / 26	34 / 13	31 / 21	44
DP	5 / 5	32 / 26	28 / 15	57 / 54	30
MP	28 / 44	22 / 31	19 / 17	5 / 8	19
RT	7 / 27	2 / 6	19 / 40	7 / 27	7
	30	28	16	26	

Figure 11.7 - Interaction among policies, approaches, and results.

example, in (a) the cell P_1-Dir shows that 100% of the P_1 models were analyzed directly while 33% of the direct analyses were performed on P_1 models. The figures to the right of each row show the percent of total effort that the row designator represents while the figures at the bottom of each column show the percent total effort of the column designator. Again in (a), observing the first row and column totals, we see that 16% of the sample involved research on P_1 policies, while in 44% of the cases the direct approach was used. Figure 11.7 confirms our analyses so far. We see from the matrices the great effort in DP analyses of P_n policies and MP analyses of C policies. The Dir analyses for P_n, P_∞ and C policies represent, for the most part, approximate and heuristic treatments.

In viewing Figure 11.7(b) it is interesting to observe that most of the analytical results (Column 1) appear for P_∞ and C. This is again due to the approximate-heuristic effort. Most of the characterization effort was put on P_n and P_∞, and the same holds true for general results. Numerical results were obtained mostly on P_n and C. The former were obtained via DP. The latter were obtained via numerical search procedures for trigger point and order quantity values after the MP (queuing) analysis with a superimposed cost function yielded an objective function of these decision variables. In considering combined a and n, C policies show up best. From a realistic point of view one should keep in mind that these often require some restrictive assumptions such as Poisson or compound Poisson demand, one-for-one ordering, no queuing in the order-filling process (ample server queuing model), and so on.

Finally, viewing Figure 11.7(c) by columns, we see the greatest possibility for yielding analytical results was a direct analysis, followed by MP. For numerical results, again the direct approach was highest, followed next by DP, and then MP. Combining a and n, direct is by far the highest followed next by MP and then DP. In observing row percentages, the

direct approach was quite successful at yielding a or n results but, of course, was used either in relatively simple cases (such as P_1 policies) or in heuristic or approximate treatments. The MP approach also produced, in half the cases on which it was used, a or n results, again mostly coming from the queuing type analyses on C policies.

We now summarize this section as follows:

1. There does not seem to be any indication of a decrease in effort in inventory research.

2. The direction of the research is changing toward heuristic and approximate models which can yield computable results.

3. A greater effort appears to be focusing on C policies, although P policies thus far have received the most attention.

4. Multi-item and multi-echelon models appear to be gaining of late in their share of attention. Much of the recent effort appears to be concentrated on workable approximate and/or heuristic models.

5. Treatment of uncertainty still occupies a small portion of the effort but this area also seems to be getting more attention of late.

6. Computational results are generally produced for P_n policies by DP and for C policies by MP. A direct probability approach (Dir) can be used to obtain computational results for heuristic and approximate models.

11.3 Survey of Practice

In Section 11.2 we have presented a survey of stochastic inventory theory from its "beginning" in the early 1950s to the present day. In this section we survey the application of inventory control techniques in practice, based mainly on situations found in the armed forces.

In order to describe the applications of inventory

control techniques in the armed services, it is necessary first to describe a military supply system. The Navy supply system illustrates the general structure. Inventory control is exercised at many levels and inventory stocks exist at many levels. The system inventory managers are called inventory control points (ICPs), national ICPs, or supply centers. Within a service all items carried in the supply system are the responsibility of one and only one ICP, which may or may not physically have an inventory. At the first lower echelon are stock points which warehouse inventory stocks and may or may not be inventory managers. For the Navy at least, a further echelon exists in the mobile logistics support force (MLSF) ships, which service operating fleet units. Finally, there are the operating forces, ships, squadrons, and battalions.

The structure of the Navy supply system is illustrated in Figure 11.8. There is a wholesale system and a retail system. Wholesale material is Navy-owned and managed by the ICPs. Retail material is managed by the Defense Supply Agency (DSA) and controlled within the Navy by the Fleet Material Support Office (FMSO). FMSO control is exercised through budget apportionment and inventory control policy specification to the stock points that carry DSA material. A further dimension of inventory control in military supply systems is Congressional funding by material groups and fiscal year.

To summarize, inventory control is exercised at the following levels.

1. ICPs--"when to buy and how much" decisions for wholesale material, and how to distribute order quantities to reporting stock points.

2. Stock points--"when to buy and how much" decisions for retail and locally procured material.

3. MLSF and tenders--construction of load lists, range and depth decisions.

4. Combatant ships--what to carry and in what depth.

Prior to approximately 1960, most "when to buy and how much" decisions were based on the simple "three and

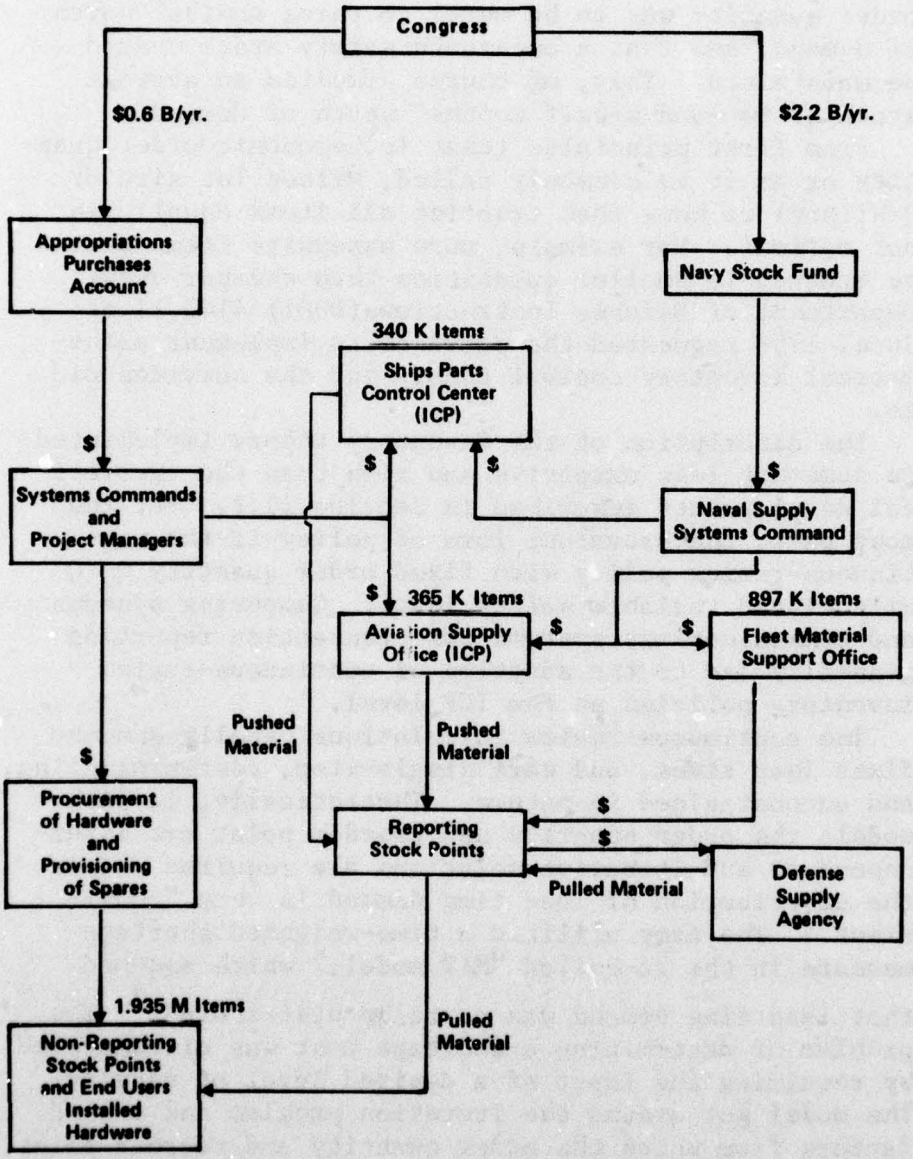


Figure 11.8 - Navy supply system.

one" rule. This rule stated that for all items, the order quantity was to be equal to three months' worth of demand, and that a one-month safety stock should be maintained. This, of course, implies an average stock of two-and-a-half months' worth of demand.

From first principles (that is, economic order quantity or as it is commonly called, Wilson lot size or Q Wilson) we know that treating all items equally is not optimal. For example, more expensive items should be ordered in smaller quantities than cheaper ones. Department of Defense Instruction (DODI) 4140.11 of June, 1958 requested the services to implement mathematical inventory control theory and the services did so.

The description of the inventory theory implemented is somewhat less extensive and rich than the theoretical developments described in Section 11.2. For the most part, the prevalent form of policy is the continuous-review policy with fixed order quantity Q (Q Wilson) and variable safety level. Computing systems and communications systems for transaction reporting generally led to the adoption of continuous-review inventory policies at the ICP level.

The continuous-review formulations usually assumed fixed lead times, and were single-item, cost-minimizing, and unconstrained in nature. Theoretically, in such models the order quantity and reorder point are interdependent and iterative solutions are required unless the distribution of lead time demand is very "convenient." The Army utilized a time-weighted shortage measure in the so-called "MIT model," which assumed that lead time demand was normally distributed.[3] The problem of determining a shortage cost was circumvented by requiring the input of a desired level of service. The model got around the iteration problem and tabled factors from which the order quantity and reorder point could easily be determined as a function of item characteristics and the desired service level. It was an

[3]Deemer, R. L. and Hoekstra, D. (1968). Improvement of M.I.T. non-repairable model. U. S. Army Logistics Management Center, Final Report AD 670977, Fort Lee, Virginia (April).

unconstrained, single-item model.

The Defense Supply Agency used a continuous-review policy with the order quantity determined independently from the reorder point. The Wilson Q was the order quantity. The reorder point was then determined as a function of Q and other factors in the standard continuous-review formulation. Details of the procedure and costs and other factors differed in each of DSA's Supply Centers.

Navy applications will now be described in more detail. At the ICP level the continuous-review model was adapted to a large multi-item, budget-constrained environment by using fixed order quantities and variable safety levels. The setting of safety levels is accomplished by substituting a material class multiplier for the individual item shortage costs, and then manipulating the multiplier until policies are produced that are fundable under the material class fiscal procurement budget. This procedure's greatest fault is that the same shortage cost is imputed for all items within a material class. If a $10 shortage cost is applied to a 10¢ item and a $10,000 item, all other things being equal, the cheaper item gets all the protection.

At the stock point level, policies for retail items are continuous-review policies with a variable order quantity in terms of months of demand and a variable safety level. For stock points, the order quantity and reorder point determinations are linked through the requirements that average investment must not exceed two-and-a-half months' worth of demand. These policies are determined in the Variable Operating and Safety Level (VOSL) program.

At the shipboard level significant (dollar-value) stocks of material are maintained. Every part installed in every assembly or equipment on board is a candidate for spare part stockage. From this large universe of candidate items, decisions must be made as to what to stock and in what depth. The criterion in this problem comes from the desire to make an individual ship self-sufficient for a specified operating period with a certain probability. The algorithm for this problem is the Fleet Logistic Support Improvement Program (FLSIP) and the shipboard stock produced is called a coordinated

shipboard allowance list, or COSAL.

As indicated by Figure 11.9, the procedure does not represent an optimization algorithm. For demand-based items the depth of stock is determined from the required operating period and probability of sufficiency desired, and from the demand probability distribution assumed. A Poisson demand distribution is assumed for all demand-based items. The CNO-directed criterion specified, prior to fiscal year 1974, 90% probability of meeting all demands for 90 days. The vast majority of COSAL items are replenished on a one-for-one basis.

Relatively few COSAL items are identified for selective item management (SIM). Such items have sufficient demand that they have continuous-review, order-quantity, reorder-point policies. In approximate magnitudes a ship with a 25,000-item COSAL may have but 300 or so SIM items. The policies for SIM items are determined from table lookups which depend on item cost and demand characteristics.

In summary, then, at the system level and stock point level the inventory models implemented have been adaptations of continuous-review policies. In most cases a fixed order quantity and variable safety level are employed. Exponential smoothing is employed in forecasting demand parameters. We also note that variance is seldom kept or forecasted directly. One scheme is not to collect and maintain variance data but simply to estimate them using an analytic function dependent on mean demand and item characteristics such as unit cost. Another popular alternative is to estimate variance from mean absolute deviation (MAD) using the relation that $\sigma = 1.25$ MAD . This relation is correct for the normal distribution and approximate for other distributions. The most popular distributions of lead time demand are the normal, Poisson, negative binomial, and an exponential as fitted to the right tail of the normal.

Shifting attention to non-military applications, we note that the IBM OS/360 inventory control application for the civilian sector is quite similar to the descriptions above. The IBM inventory control package employs a fixed-order quantity (Q Wilson or months of supply) with a variable safety level and exponential smoothing for demand forecasting. The question of determining shortage costs is circumvented by requiring the input of

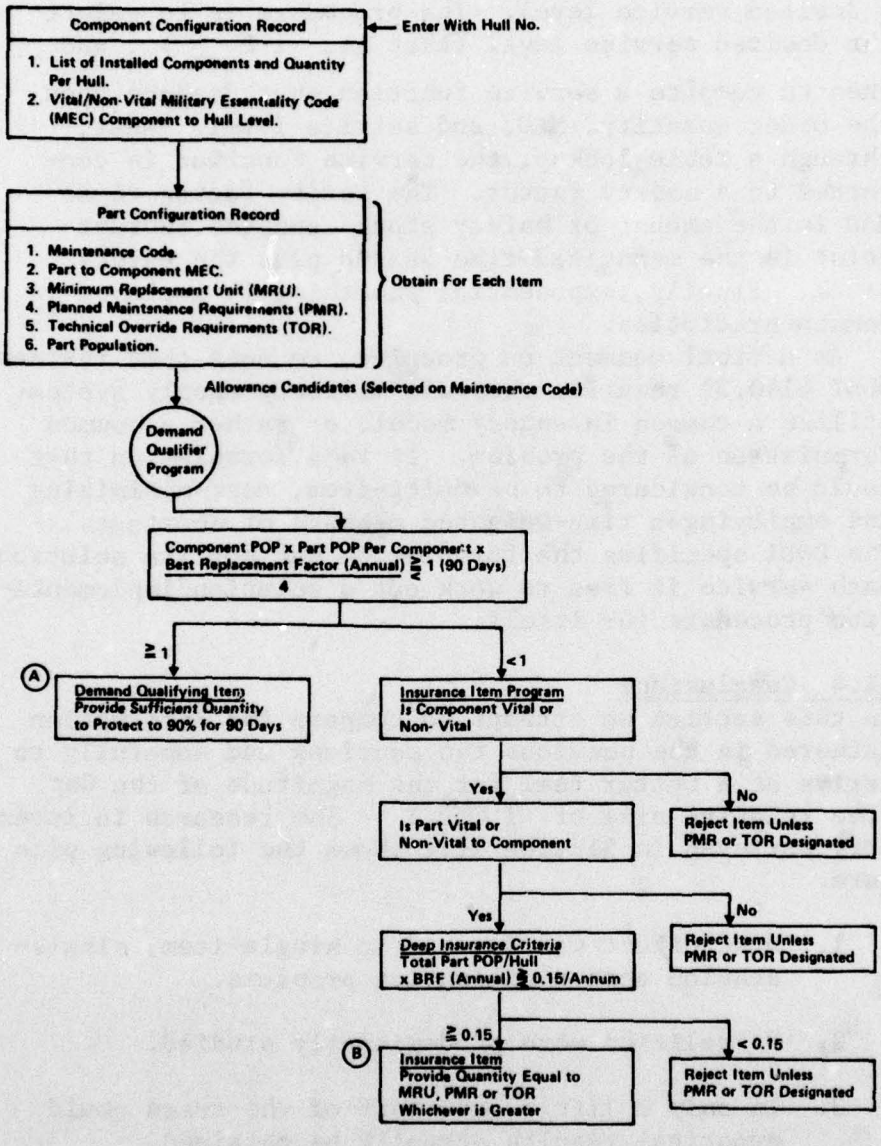


Figure 11.9 - Computation of Fleet Logistics Support Improvement Program (FLSIP) COSALS.

a desired service level. The procedure is to select the desired service level (that is, $1-P_{out}$), and then to compute a service function that depends upon the order quantity, MAD, and service level. Next, through a table lookup, the service function is converted to a safety factor. The safety factor times MAD is the amount of safety stock, and the reorder point is the mean lead time demand plus the safety stock. Finally, exponential smoothing is employed in demand prediction.

As a final comment on practice, we note that the new DODI 4140.39 requires that all military supply systems utilize a common inventory model, or rather a common formulation of the problem. It is a formulation that could be considered to be multi-item, cost-minimizing and employing a time-weighted measure of shortages. The DODI specifies the formulation but not its solution. Each service is free to work out a solution/implementation procedure for itself.

11.4 Conclusions

In this section we attempt to compare the information gathered in the previous two sections and hopefully to arrive at a better feel for the magnitude of the Gap (the relative size of $T \cap R$). The research in inventory surveyed in Section 11.2 shows the following picture.

1. Most effort was devoted to single-item, single-echelon cost minimization problems.

2. P policies were predominantly studied.

3. In only a little over half of the cases could numerical results actually be obtained.

4. Very little of the research effort took into account uncertainty (forecasting of demand). Ninety percent of all models were decision-making under risk (probability distributions assumed completely known).

The practice surveyed in Section 11.3 indicates the

following.

1. Predominantly, approximate continuous-review models are in use.

2. Constraints on service level, budget, and so on, are used as secondary criteria in conjunction with the minimization of costs.

3. Forecasting is always necessary and generally is accomplished by exponential smoothing.

4. Problems are really multi-product and multi-echelon in nature.

Comparing these two lists, there is evidence of a considerable Gap between theory and practice. However, as we pointed out in Section 11.2, there seems to be a change in the research direction toward approximate and heuristic models, which easily yield numbers, more emphasis on C policies, and greater concern over multi-product and multi-echelon models. Also, with the advances in nonlinear programming, more consideration appears to be given to constrained optimization (see, for example, Schrady and Choe (1971)).

With respect to problems of forecasting, there has been relatively little effort shown in the survey on research (recall that approximately 11% of the sample considered problems of uncertainty); however, even here a slight trend toward this end may be present. Of the 11% of the papers that considered uncertainty problems at all, about 44% were published after 1969. Thus, while it appears to us that a sizable Gap does exist (putting a numerical value on it would be highly subjective and we refrain from doing so here), it also appears to us that the change in the direction of research effort is toward a narrowing of the Gap.

We conclude this survey by noting several supply problems which perhaps require not so much a new model or technique but the careful orchestration of analytic tools and techniques. The first of these concerns the general area of data collection and forecasting for inventory control. If the data are deficient, the model, be it for inventory control or provisioning, is

only a garbage processor and management still knows nothing about how to proceed--though, unfortunately, after processing the data through the model, data deficiencies are obscured. The transaction reporting systems that most supply systems use do much to aggregate data in peculiar ways. Some examples of the kind of demand data seen at the system or ICP level are given in Figures 11.10-.12.[4]

Another significant problem area is that of supply system response time. The ability of a supply system to respond to the need for material is obviously a function of inventory stocking rules such as whether material is on hand or a stock-out condition exists. However, response time also depends on decisions about whether an item will be stocked at a given echelon at all (Pinkus et al. (1973) is one of the few theoretical attempts to treat this problem). Further, it depends on the procedures used in requisition processing, the degree of automation, the transportation modes used, and the level of resources devoted to all these functions. Intensive management and high levels of stock in the system can be traded off; today's funding climate does not allow both, and so this trade-off problem is very important.[5]

Finally, there is a need in large systems of any kind to assure that goals are structured for individual activities of the system, so that the system objectives are achieved. For stock points the traditional measure of effectiveness has been net supply effectiveness. This is the percent of requisitions filled from stock on hand for items that the stock point is "supposed" to carry. With the marginal funding in the years since the buildup for the conflict in Southeast Asia, we now have stock points periodically redefining their stocking criteria. This is apparently being done without any overall guidance or control by the system. The stocking criteria are being made more stringent so that the

---

[4]For an example of relatively recent work on this "lumpy demand" problem, see Silver (1970).

[5]Prichard, J. W. et al. (1973). Ships supply support study. Department of the Navy, Washington. (15 June).

AD-A050 798 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 15/5
MODERN TRENDS IN LOGISTICS RESEARCH. PROCEEDINGS OF A CONFERENC--ETC(U)
1976 W H MARLOW N00014-75-C-0729
UNCLASSIFIED NL

4 OF 5
AD A050798



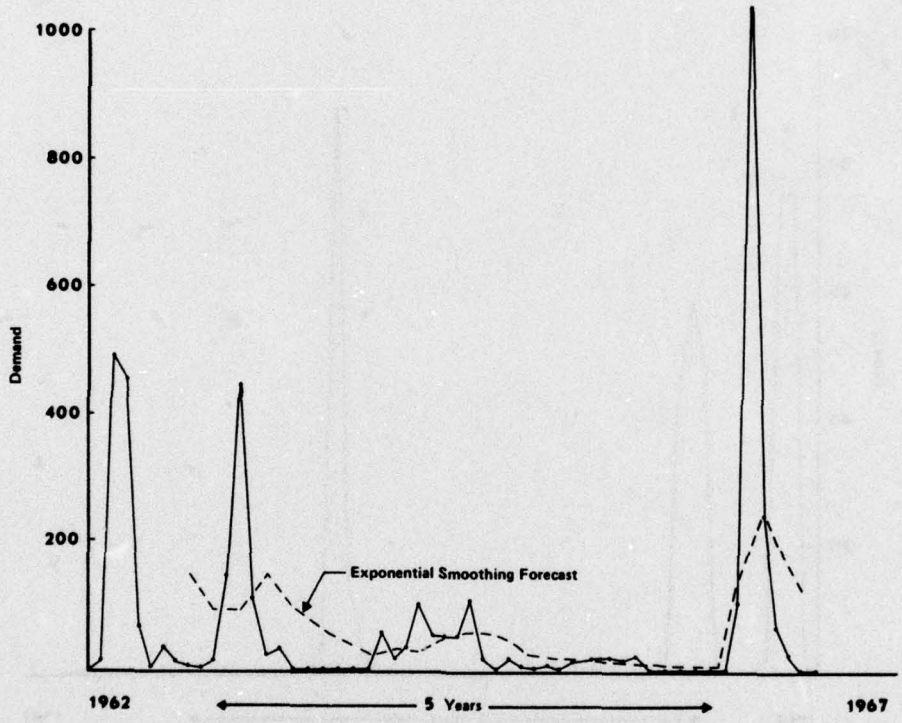


Figure 11.10 - Five years of monthly system demand observations, item unit cost = $0.30, sample item 273.

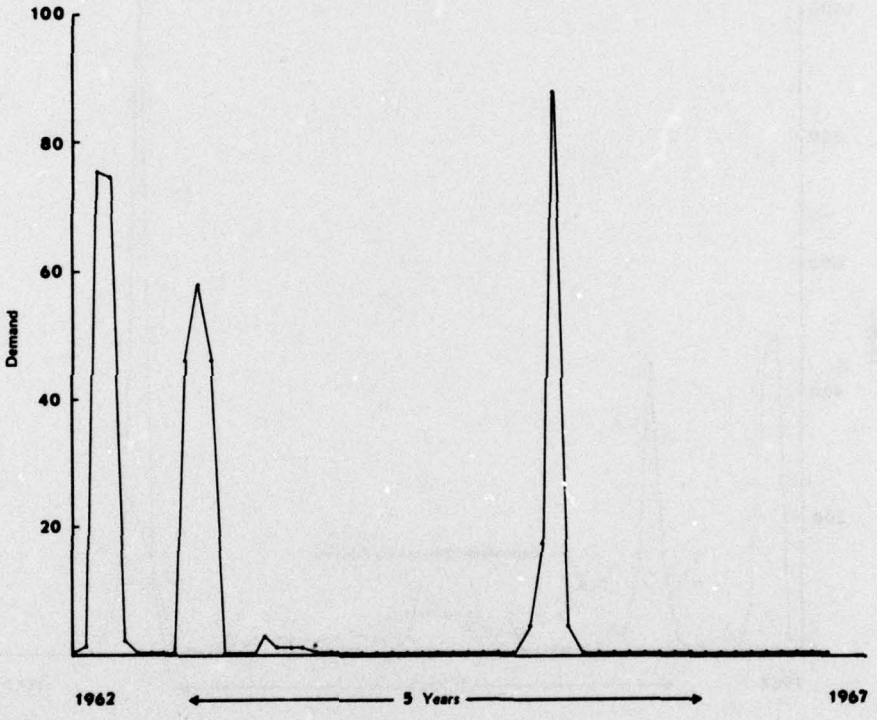


Figure 11.11 - Five years of monthly system demand observations, item unit cost = $23.00, sample item 488.

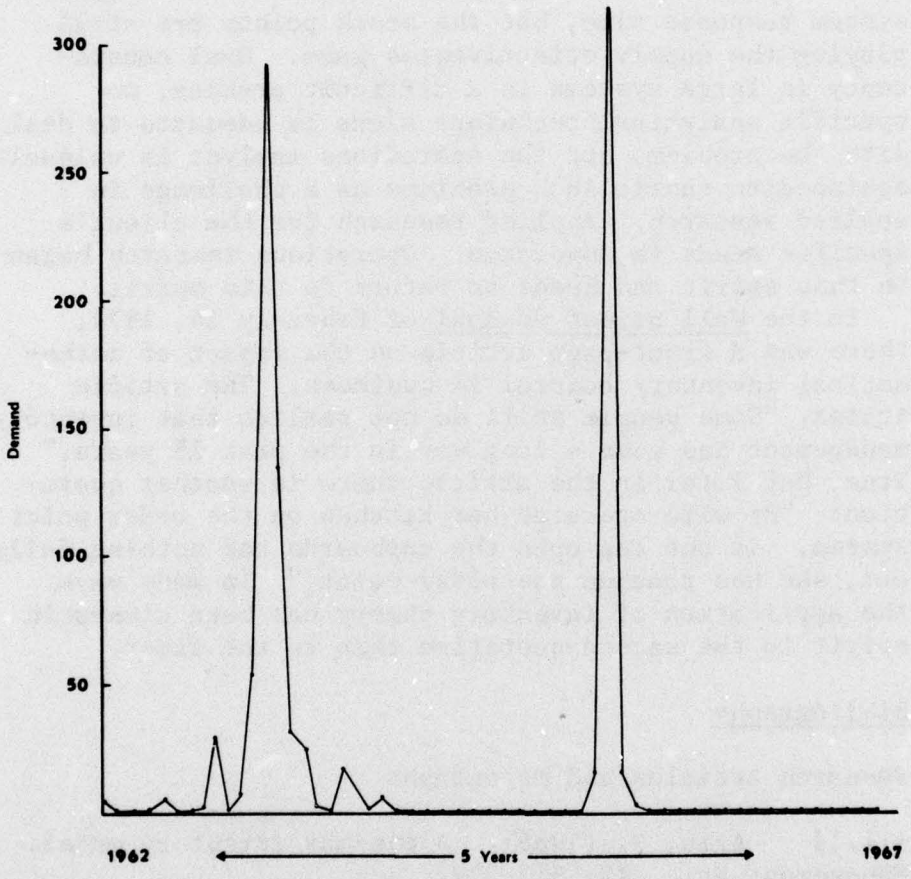


Figure 11.12- Five years of monthly system demand observations, item unit cost = $3.50, sample item 738.

available dollars are spread over fewer items and supply effectiveness can be maintained. By doing this, the stock point's report card looks good but the impact on response time is unclear and probably detrimental. At the system level the new measure of effectiveness is system response time, but the stock points are still playing the supply effectiveness game. Goal consistency in large systems is a difficult problem; no specific analytical technique alone is adequate to deal with the problem, but the operations analyst is uniquely equipped to tackle such problems as a challenge in applied research. Applied research for the client's specific needs is advocated. Operations research began in this spirit and needs to return to this spirit.

In the Wall Street Journal of February 14, 1974, there was a front-page article on the impact of mathematical inventory control in business. The article stated, "Some people still do not realize that inventory mat[illegible] has come a long way in the past 15 years." True, [illegible]later in the article there is another quotation: "My wife operates her kitchen on the order point system. If she can open the cupboards and nothing falls out, she has reached the order point." In many ways the application of inventory theory has been closer in spirit to the second quotation than to the first.

Bibliography

Research Articles and Monographs

[11.1] Agin, N. (1966). A min-max inventory model. Management Sci. 12 517-529.

[11.2] Allen, S. G. (1958). Redistribution of total stock over several user locations. Naval Res. Logist. Quart. 15 337-345.

[11.3] Allen, S. G. (1961). A redistribution model with set-up charge. Management Sci. 8 99-108.

[11.4] Allen, S. G. (1962). Computation for the redistribution model with set-up charge. Management Sci. 8 482-489.

[11.5] Arrow, K. J., T. Harris, and J. Marschak (1951). Optimal inventory policy. Econometrica 19 250-272.

[11.6] Balintfy, J. L. (1964). On a basic class of multi-item inventory problems. Management Sci. 10 287-297.

[11.7] Barankin, E. W. (1961). A delivery-lag inventory model with an emergency provision (the single period case). Naval Res. Logist. Quart. 8 285-311.

[11.8] Bather, J. A. (1966). A continuous time inventory model. J. Appl. Probability 3 538-549.

[11.9] Beckmann, M. (1965). An inventory model for arbitrary interval and quantity distributions of demand. in A. F. Veinott, Jr. (ed.) Mathematical Studies in Management Science. Macmillan. 422-444.

[11.10] Beckmann, M., and R. Muth (1956). An inventory policy for a case of lagged delivery. Management Sci. 2 145-155.

[11.11] Beckmann, M., and R. Muth (1958). On the two-bin inventory policy: an application of the Arrow-Harris-Marschak model. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 210-219.

[11.12] Beesack, P. R. (1967). A finite horizon dynamic model with a stockout constraint. Management Sci. 13 618-630.

[11.13] Bellman, R., I. Glicksberg, and O. Gross (1955). On the optimal inventory equation. Management Sci. 2 83-109.

[11.14] Bessler, S. A., and A. F. Veinott, Jr. (1966). Optimal policy for a dynamic multi-echelon model. Naval Res. Logist. Quart. 13 355-389.

[11.15] Bessler, S. A., and P. W. Zehna (1968). An

application of servomechanisms to inventory. Naval Res. Logist. Quart. 15 157-168.

[11.16] Boylan, E. S. (1964). Multiple (s,S) policies. Econometrica 32 399-409.

[11.17] Boylan, E. S. (1967). Multiple (s,S) policies and the n-period inventory problem. Management Sci. 14 196-204.

[11.18] Brown, G. F., Jr., T. M. Corcoran, and R. M. Lloyd (1971). Inventory models with forecasting and dependent demands. Management Sci. 17 498-499.

[11.19] Brown, G. F., Jr., and W. F. Rogers (1973). A Bayesian approach to demand estimation and inventory provisioning. Naval Res. Logist. Quart. 20 607-624.

[11.20] Chern, Ho Chung-Mei (1974). A multi-product joint ordering model with dependent set-up cost. Management Sci. 20 1081-1091.

[11.21] Clark, A. J. (1960). The use of simulation to evaluate a multi-echelon, dynamic inventory model. Naval Res. Logist. Quart. 7 429-445.

[11.22] Clark, A. J., and H. Scarf (1960). Optimal policies for a multi-echelon inventory problem. Management Sci. 6 475-490.

[11.23] Clark, A. J., and H. Scarf (1962). Approximate solutions to a simple multi-echelon inventory problem. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in Applied Probability and Management Science. Stanford University Press, Stanford, California. 88-100.

[11.24] Connors, M. M., and W. I. Zangwill (1971). Cost minimization in networks with discrete stochastic requirements. Operations Res. 19 794-821.

[11.25] Crowston, W. B., W. H. Hausman, and W. R. Kampe (1973). Multistage production for stochastic seasonal demand. Management Sci. 19 924-935.

[11.26] Daniel, K. (1963). A delivery-lag inventory model with emergency. in H. Scarf, D. M. Gilford, and M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 32-46.

[11.27] Denicoff, M., J. Fennell, S. E. Haber, W. H. Marlow, F. W. Segel, and H. Solomon (1964). The Polaris military essentiality system. Naval Res. Logist. Quart. 11 235-257.

[11.28] Denicoff, M., J. Fennell, S. E. Haber, W. H. Marlow, and H. Solomon (1964). A Polaris logistics model. Naval Res. Logist. Quart. 11 259-272.

[11.29] Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1952a). The inventory problem I: case of known distribution of demand. Econometrica 20 187-227.

[11.30] Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1952b). The inventory problem II: case of unknown distributions of demand. Econometrica 20 450-466.

[11.31] Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1953). On the optimal character of the (s,S) policy in inventory theory. Econometrica 21 586-596.

[11.32] Eilon, S., and J. Elmaleh (1968). An evaluation of alternate inventory control policies. Internat. J. Production Res. 7.

[11.33] Eilon, S., and J. Elmaleh (1970). Adaptive limits in inventory control. Management Sci. 16 533-548.

[11.34] Evans, R. V. (1967). Inventory control of a multi-product system with a limited production resource. Naval Res. Logist. Quart. 14 173-184.

[11.35] Evans, R. V. (1968). Sales and restocking policies in a single-item inventory system. Management Sci. 14 463-472.

[11.36] Feeney, G. J., and C. C. Sherbrooke (1966). The (s-1,s) inventory policy under compound Poisson

demand. Management Sci. 12 391-411.

[11.37] Fromowitz, S. (1965). A class of one-period inventory models. Operations Res. 13 779-799.

[11.38] Fukuda, Y. (1961). Optimal disposal policies. Naval Res. Logist. Quart. 8 221-227.

[11.39] Gallagher, D. J. (1969). Two periodic review inventory models with backorders and stuttering Poisson demands. American Institute of Industrial Engineers Transactions 1 164-171.

[11.40] Galliher, H. P., P. M. Morse, and M. Simond (1959). Dynamics of two classes of continuous review inventory systems. Operations Res. 7 362-384.

[11.41] Gaver, D. P., Jr. (1959). On base-stock level inventory control. Operations Res. 7 689-703.

[11.42] Goyal, S. K. (1973). Lot-size scheduling on a single machine for stochastic demand. Management Sci. 19 1322-1325.

[11.43] Greenberg, H. (1964). Time-dependent solutions to the (s,S) inventory problem. Operations Res. 12 725-735.

[11.44] Gross, D. (1963). Centralized inventory control in multilocation supply systems. in H. Scarf, D. M. Gilford, and M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 47-84.

[11.45] Gross, D., and C. M. Harris (1971). On one-for-one ordering inventory policies with state-dependent lead times. Operations Res. 19 735-760.

[11.46] Gross, D., and C. M. Harris (1973). Continuous review (s,S) inventory models with state-dependent lead times. Management Sci. 19 567-574.

[11.47] Gross, D., and A. Soriano (1969). The effect of reducing lead times on inventory levels--simulation

analysis. Management Sci. 16 B-61 - B-76.

[11.48] Gross, D., and A. Soriano (1972). On the economic application of airlift to product distribution and its impact on inventory levels. Naval Res. Logist. Quart. 19 501-507.

[11.49] Haber, S. E. (1971). Simulation of multi-echelon macro-inventory policies. Naval Res. Logist. Quart. 18 119-134.

[11.50] Haber, S. E., and R. Sitgreaves (1970). Methodology for estimating expected usage of repair parts with application to parts with no usage history. Naval Res. Logist. Quart. 17 535-546.

[11.51] Haber, S. E., and R. Sitgreaves (1972). A unified model for demand prediction in the context of provisioning and replenishment. Naval Res. Logist. Quart. 19 29-42.

[11.52] Haber, S. E., R. Sitgreaves, and H. Solomon (1969). A demand prediction technique for items in military inventory systems. Naval Res. Logist. Quart. 16 297-308.

[11.53] Hadley, G., and T. M. Whitin (1961a). A family of inventory models. Management Sci. 7 351-371.

[11.54] Hadley, G., and T. M. Whitin (1961b). A model for procurement allocation and redistribution for low demand items. Naval Res. Logist. Quart. 8 395-414.

[11.55] Hadley, G., and T. M. Whitin (1963). An inventory-transportation model with N locations. in H. Scarf, D. M. Gilford, and M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 116-142.

[11.56] Hanssmann, F. (1959). Optimal inventory location and control in production and distribution networks. Operations Res. 7 483-498.

[11.57] Harris, C. M. (1971). Some statistical results for inventory models with state-dependent lead times. Colloquia Mathematica Societatis János Bolyai 7. Inventory Control and Water Storage, Győr, Hungary. 105-120.

[11.58] Harris, F. (1915). Operations and Cost (Management Series) 48-52. A. W. Shaw Company, Chicago.

[11.59] Hartfiel, D. J., and G. L. Curry (1974). An algorithm for optimal inventory policies for systems with joint set-up costs. Management Sci. 20 1175-1177.

[11.60] Hartung, P. H. (1973). A simple style goods inventory model. Management Sci. 19 1452-1458.

[11.61] Hausman, W. H. (1969). Minimizing customer-line items backordered in inventory control. Management Sci. 15 B-628 - B-634.

[11.62] Hausman, W. H., and R. Peterson (1972). Multiproduct production scheduling for style goods with limited capacity forecast revision and terminal delivery. Management Sci. 18 370-383.

[11.63] Hausman, W. H., and R. S. Sides (1973). Mail-order demands for style goods: theory and data analysis. Management Sci. 20 191-202.

[11.64] Hausman, W. H., and L. J. Thomas (1972). Inventory control with probabilistic demand and periodic withdrawals. Management Sci. 18 265-275.

[11.65] Hayes, R. H. (1969). Statistical estimation problems in inventory control. Management Sci. 15 686-701.

[11.66] Ho, C. M. (1970). A note on the calculation of expected time-weighted backorders over a given interval. Naval Res. Logist. Quart. 17 555-559.

[11.67] Hochstaedter, D. (1970). An approximation of cost-function for multi-echelon inventory models.

Management Sci. 16 716-727.

[11.68] Holt, C. C., F. Modigliani, and J. F. Muth (1956). Derivation of a linear decision rule for production and employment. Management Sci. 2 159-177.

[11.69] Hunt, J. (1965). Balancing accuracy and simplicity in determining reorder points. Management Sci. 12 B-94 - B-103.

[11.70] Hurter, A. P., and F. C. Kaminsky (1967). An application of regenerative stochastic processes to a problem in inventory control. Operations Res. 15 467-472.

[11.71] Iglehart, D. L. (1963a). Dynamic programming and stationary analysis of inventory problems. in H. Scarf, D. M. Gilford, M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 1-31.

[11.72] Iglehart, D. L. (1963b). Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. Management Sci. 9 254-267.

[11.73] Iglehart, D. L. (1964). The dynamic inventory problem with unknown demand distribution. Management Sci. 10 429-440.

[11.74] Iglehart, D. L., and S. Karlin (1962). Optimal policy for dynamic inventory process with non-stationary stochastic demands. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in Applied Probability and Management Science. Stanford University Press, Stanford, California. 127-147.

[11.75] Iglehart, D. L., and A. Lalchandani (1967). An allocation model. SIAM J. Appl. Math. 15 303-323.

[11.76] Iglehart, D. L., and R. C. Morey (1971). Optimal policies for a multi-echelon inventory system with demand forecasts. Naval Res. Logist. Quart. 18 115-118.

[11.77] Iglehart, D. L., and R. C. Morey (1972). Inventory systems with imperfect asset information. Management Sci. 18 B-388 - B-394.

[11.78] Ignall, E. (1969). Optimal continuous review policies for two-product inventory systems with joint set-up costs. Management Sci. 15 278-283.

[11.79] Ignall, E., and A. F. Veinott, Jr. (1969). Optimality of myopic inventory policies for several substitute products. Management Sci. 15 284-304.

[11.80] Johnson, E. L. (1967). Optimality and computation of (σ,S) policies in multi-item infinite horizon inventory problem. Management Sci. 13 475-491.

[11.81] Johnson, E. L. (1968). On (s,S) policies. Management Sci. 15 80-101.

[11.82] Kaplan, A. J. (1969). Stock rationing. Management Sci. 15 260-267.

[11.83] Kaplan, A. J. (1973). A stock redistribution model. Naval Res. Logist. Quart. 20 231-239.

[11.84] Kaplan, R. S. (1970). A dynamic inventory model with stochastic lead times. Management Sci. 16 491-507.

[11.85] Karlin, S. (1958a). One-stage inventory models with uncertainty. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 109-134.

[11.86] Karlin, S. (1958b). Optimal inventory policy for the Arrow-Harris-Marschak dynamic model. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 135-154.

[11.87] Karlin, S. (1958c). Steady-state solutions. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies

in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 223-269.

[11.88] Karlin, S. (1958d). The application of renewal theory to the study of inventory policies. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 270-297.

[11.89] Karlin, S. (1960). Dynamic inventory policy with varying stochastic demands. Management Sci. 6 231-258.

[11.90] Karlin, S., and H. Scarf (1958a). Inventory models of the Arrow-Harris-Marschak type with time lag. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 155-178.

[11.91] Karlin, S., and H. Scarf (1958b). Inventory models and related stochastic processes. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 319-336.

[11.92] Karr, H. W. (1958). A method of estimating spare-part essentiality. Naval Res. Logist. Quart. 5 29-42.

[11.93] Karush, W. (1957). A queuing model for an inventory problem. Operations Res. 5 693-703.

[11.94] Kasugai, H., and T. Kasegai (1960). Characteristics of dynamic maximin ordering policy. J. Operations Res. Soc. Japan 3 11-26.

[11.95] Kasugai, H., and T. Kasegai (1961). Note on minimax regret ordering policy. Static and dynamic solutions and a comparison to maximin policy. J. Operations Res. Soc. Japan 3 155-169.

[11.96] Laderman, J., S. B. Littauer, and L. Weiss (1953). The inventory problem. J. Amer. Statist. Assoc. 48 717-752.

[11.97] Levy, J. (1958). Loss resulting from the use of incorrect data in computing an optimal inventory policy. Naval Res. Logist. Quart. 5 75-82.

[11.98] Levy, J. (1959). Further notes on the loss resulting from the use of incorrect data in computing an optimal inventory policy. Naval Res. Logist. Quart. 6 25-32.

[11.99] Love, R. F. (1967a). Inventory policy when order interarrival times are gamma distributed and reorder lead times are exponentially distributed. CORS J. 5 27-34.

[11.100] Love, R. F. (1967b). A two-station stochastic inventory model with exact methods of computing optimal policies. Naval Res. Logist. Quart. 14 185-217.

[11.101] Maher, M. J., J. C. Gittens, and R. W. Morgan (1973). An analysis of a multi-line reorder system using a can-order policy. Management Sci. 19 800-808.

[11.102] Manne, A. S. (1960). Linear programming and sequential decisions. Management Sci. 6 259-267.

[11.103] Montgomery, D. C., M. S. Bazaraa, and A. K. Keswani (1973). Inventory models with a mixture of backorders and lost sales. Naval Res. Logist. Quart. 20 255-263.

[11.104] Morey, R. C. (1970). Inventory systems with imperfect demand information. Naval Res. Logist. Quart. 17 287-295.

[11.105] Morse, P. M. (1958). Chapter 10 of Queues, Inventories and Maintenance. Wiley.

[11.106] Morse, P. M. (1959). Solutions of a class of discrete time inventory problems. Operations Res. 7

67-78.

[11.107] Morton, T. E. (1969). Bounds on the solution of the lagged optimum inventory equation with no demand backlogging and proportional costs. SIAM Rev. 11 572-596.

[11.108] Morton, T. E. (1971). The near-myopic nature of the lagged proportional-cost inventory problem with lost sales. Operations Res. 19 1708-1715.

[11.109] Muckstadt, J. A. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory system. Management Sci. 20 472-481.

[11.110] Neuts, M. (1964). An inventory model with an optional time lag. SIAM J. Appl. Math. 12 179-185.

[11.111] Oral, M., M. S. Salvador, A. Reisman, and B. V. Dean (1972). On the evaluation of shortage costs in inventory control of finished goods. Management Sci. 18 B-344 - B-351.

[11.112] Pinkus, C. E., D. Gross, and R. M. Soland (1973). Optimal design of multiactivity, multifacility systems by branch and bound. Operations Res. 21 270-283.

[11.113] Porteus, E. L. (1971). On the optimality of generalized (s,S) policies. Management Sci. 17 411-426.

[11.114] Porteus, E. L. (1972). The optimality of generalized (s,S) policies under uniform demand densities. Management Sci. 18 644-646.

[11.115] Posner, M. J. M., and B. Yansouni (1972). A class of inventory models with customer impatience. Naval Res. Logist. Quart. 19 483-492.

[11.116] Ravindran, A. (1972). Management of seasonal style-goods inventories. Operations Res. 20 265-275.

[11.117] Roberts, D. M. (1962). Approximations to optimal policies in a dynamic inventory model. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in Applied Probability and Management Science. Stanford University Press, Stanford, California. 207-229.

[11.118] Rose, M. (1972). The (S-1,S) inventory model with arbitrary backordered demand and constant delivery time. Operations Res. 20 1020-1032.

[11.119] Scarf, H. (1958a). A min-max solution of an inventory problem. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 201-209.

[11.120] Scarf, H. (1958b). Stationary operating characteristics of an inventory model with time lag. in K. J. Arrow, S. Karlin, and H. Scarf (eds.) Studies in the Mathematical Theory of Inventory and Production. Stanford University Press, Stanford, California. 298-318.

[11.121] Scarf, H. (1959). Bayes solutions of the statistical inventory problem. Ann. Math. Statist. 30 490-508.

[11.122] Scarf, H. (1960a). Some remarks on Bayes solutions to the inventory problem. Naval Res. Logist. Quart. 7 591-596.

[11.123] Scarf, H. (1960b). The optimality of (S,s) policies in the dynamic inventory problem. in K. J. Arrow, S. Karlin, and P. Suppes (eds.) Mathematical Methods in the Social Sciences, 1959. Stanford University Press, Stanford, California. 196-202.

[11.124] Schaack, J. P., and E. A. Silver (1972). A procedure involving simulation for selecting the control variables of an (S,c,s) joint ordering strategy. INFOR-Canad. J. Operational Res. and Information Processing 10 154-170.

[11.125] Schrady, D. A., and U. C. Choe (1971). Models for multi-item continuous review inventory policies subject to constraints. Naval Res. Logist. Quart. 18 451-463.

[11.126] Sherbrooke, C. C. (1968). METRIC: a multi-echelon technique for recoverable item control. Operations Res. 16 122-141.

[11.127] Silver, E. A. (1965). Some characteristics of a special joint-order inventory model. Operations Res. 13 319-327.

[11.128] Silver, E. A. (1970). Some ideas related to the inventory control of items having erratic demand patterns. CORS J. 8 87-100.

[11.129] Silver, E. A. (1972). Inventory allocation among an assembly and its repairable subassemblies. Naval Res. Logist. Quart. 19 261-280.

[11.130] Silver, E. A. (1973). Three ways of obtaining the average cost expression in a problem related to joint replenishment inventory control. Naval Res. Logist. Quart. 20 241-254.

[11.131] Silver, E. A., C. M. Ho, and R. L. Deemer (1971). Cost minimizing inventory control of items having a special type of erratic demand pattern. INFOR-Canad. J. Operational Res. and Information Processing 9 198-219.

[11.132] Simmons, D. M. (1972). Optimal inventory policies under a hierarchy of set-up costs. Management Sci. 18 591-599.

[11.133] Simon, R. M. (1971). Stationary properties of a two-echelon inventory model for low demand items. Operations Res. 19 761-773.

[11.134] Simpson, K. F., Jr. (1959a). In process inventories. Operations Res. 6 863-872.

[11.135] Simpson, K. F., Jr. (1959b). A theory of

allocation of stocks to warehouses. Operations Res. 7 797-805.

[11.136] Sivazlian, B. D. (1974). A continuous-review (s,S) inventory system with arbitrary interarrival distribution between unit demand. Operations Res. 22 65-71.

[11.137] Tan, F. K. (1974). Optimal policies for a multi-echelon inventory problem with periodic ordering. Management Sci. 20 1104-1111.

[11.138] Tijms, H. C. (1971). The optimality of (s,S) inventory policies in the infinite period model. Statistica Neerlandica 25 (1).

[11.139] Topkis, D. (1968). On optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes. Management Sci. 15 160-176.

[11.140] Vassian, H. J. (1955). Applications of discrete variable servo theory to inventory control. Operations Res. 3 272-282.

[11.141] Veinott, A. F., Jr. (1963). Optimal policies with nonstationary stock demands. in H. Scarf, D. M. Gilford, and M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 85-115.

[11.142] Veinott, A. F., Jr. (1965a). Optimal policy for a multi-product, dynamic, non-stationary inventory problem. Management Sci. 12 206-222.

[11.143] Veinott, A. F., Jr. (1965b). The optimal policy for batch ordering. Operations Res. 13 424-432.

[11.144] Veinott, A. F., Jr. (1965c). Optimal policy in a dynamic, single-product, non-stationary inventory model with several demand classes. Operations Res. 13 761-778.

[11.145] Veinott, A. F., Jr. (1966). On the optimality of (s,S) inventory policies. New condition and a new proof. SIAM J. Appl. Math. 14 1067-1083.

[11.146] Veinott, A. F., Jr. (1971). Least d-majorized network flows with inventory and statistical applications. Management Sci. 17 547-567.

[11.147] Veinott, A. F., Jr., and H. M. Wagner (1965). Computing optimal (s,S) inventory policies. Management Sci. 11 525-552.

[11.148] Wagner, H. M. (1962). Statistical management of inventory systems. Wiley.

[11.149] Wagner, H. M., M. O'Hagan, and B. Lundh (1965). An empirical study of exactly and approximately optimal inventory policies. Management Sci. 11 690-723.

[11.150] Wheeler, A. C. (1972). Stationary (s,S) policies for a finite horizon. Naval Res. Logist. Quart. 19 601-619.

[11.151] Whitin, T. M., and J. W. T. Youngs (1955). A method for calculating optimal inventory levels and delivery time. Naval Res. Logist. Quart. 2 157-173.

[11.152] Williams, J. F. (1974). Multi-echelon production scheduling when demand is stochastic. Management Sci. 20 1253-1263.

[11.153] Wilson, R. H., and W. A. Mueller (1926-27). A new method for stock control. Harvard Business Rev. 5 197-205.

[11.154] Wright, G. (1968). Optimal policies for a multi-product inventory system with negotiable lead times. Naval Res. Logist. Quart. 15 375-402.

[11.155] Yaspan, A. (1972). Fixed-stockout-probability order quantities with lost sales and time lag. Operations Res. 20 903-904.

[11.156] Zabel, E. (1962). A note on the optimality of (s,S) policies in inventory theory. Management Sci. 9 123-125.

[11.157] Zacks, S. (1969). Bayes sequential design of stock levels. Naval Res. Logist. Quart. 16 143-155.

[11.158] Zacks, S. (1970). A two-echelon, multi-station inventory model for Navy applications. Naval Res. Logist. Quart. 17 79-85.

[11.159] Zacks, S., and J. Fennell (1972). Bayes adaptive control of two-echelon inventory systems I: development for a special case of one-station lower echelon and Monte Carlo evaluation. Naval Res. Logist. Quart. 19 15-28.

Survey Articles

[11.160] Clark, A. J. (1972). An informal survey of multi-echelon inventory theory. Naval Res. Logist. Quart. 19 621-650.

[11.161] Hanssmann, F. (1961). A survey of inventory theory from the Operations Research viewpoint. Progress in Operations Research (R. L. Ackoff, ed.) 65-104. Wiley.

[11.162] Iglehart, D. L. (1967). Recent results in inventory theory. J. Indust. Eng. 18 48-51.

[11.163] Scarf, H. (1963). A survey of analytic techniques in inventory theory. in H. Scarf, D. M. Gilford, and M. W. Shelly (eds.) Multistage Inventory Models and Techniques. Stanford University Press, Stanford, California. 185-225.

[11.164] Veinott, A. F., Jr. (1966). The status of mathematical inventory theory. Management Sci. 12 745-777.

Books

[11.165] Hadley, G., and T. M. Whitin (1963). Analysis of Inventory Systems. Prentice-Hall.

[11.166] Howard, R. (1960). Dynamic Programming and Markov Processes. Wiley.

[11.167] Magee, J. F. (1958). Production Planning and Inventory Control. McGraw-Hill.

[11.168] Naddor, E. (1966). Inventory Systems. Wiley.

[11.169] Starr, M. K., and D. W. Miller (1962). Inventory Control Theory and Practice. Prentice-Hall.

[11.170] Whitin, T. M. (1953). The Theory of Inventory Management. Princeton University Press.

Chapter 12

PROBABILITY MODELS IN LOGISTICS*

Donald P. Gaver
Naval Postgraduate School

12.1 Introduction

The term logistics covers a wide range of special topics, even if one attempts to be restricted by Webster's definition: "that branch of the military art which embraces the details of the transport, quartering, and supply of troops." Current usage of this term certainly extends broadly to include many non-military governmental and civilian activities, so it is not surprising that probabilistic methods have been applied to only a few of the possible logistics situations that arise in our complex society. This chapter in no way aims to be a comprehensive survey, but rather points out and briefly explores certain problem types, and approaches to problems, that have been, or promise to be, of interest in the logistics area, and that seem to be approachable by probabilistic methods.

12.2 A Logistics Interpretation of the Birthday Problem

Many of us are familiar with the famous "birthday problem" that asks: "How many individuals must be present in a group in order that the chance be 50% that at least two group members have the same birthday?" This problem is discussed by Feller [12.1](1, p. 33) and is an example of an occupancy problem in classical probability. Many were also initially surprised to learn that the answer is only about 23. No doubt some of us who teach probability have been chagrined at responses of classes much larger than 23 that have admitted to no common birthdays.

Our purpose here is to show that the "birthday problem" is analogous to a problem in logistics--specifically, in supply--and, in the process, to describe a

*The author gratefully acknowledges the research support of the Office of Naval Research and the National Science Foundation Grant Number AG-476.

simple dynamic version of the usual birthday problem formulation.

A reformulation of the birthday problem is the following. Imagine that an observer stands at the corner of Hollywood and Vine (or Pennsylvania and 18th, or Columbus and Broadway, as luck dictates). He asks passersby the day of their birth, and waits until for the first time he has received exactly r ($r \geq 2$) identical replies (for example for $r = 2$, until 2 individuals have appeared with, say, May 5 birthdays; otherwise there are no matches). At this point he goes home. We are interested in the time he must wait until the experiment is complete. If passersby appear at unit rate then we might expect a match before 23 customers appear with probability about 50%, by analogy with the classical formulation.

In order to relate this experiment to a logistics situation, simply identify possible birthdays with with distinct failure-prone *elements on a ship* that goes out to sea on a mission. Of course, there will be many more than 365 such distinct elements. Let the ith such element be backstopped by $k_i - 1$ spares, permit neither interchangeability of these nor replenishment or repair of spares during the mission. Next, suppose that elements fail at random, that is, each fails independently and in accordance with a Poisson process, the ith at rate λ_i. What, then, is the probability distribution of the time until <u>some</u> element is rendered completely inoperative, owing to one failure more than k_i, for some i? Notice that if $k_i = 1$ and $\lambda_i = \lambda$ then this is equivalent to a dynamic birthday problem, in which passersby appear in Poisson manner with overall rate $\sum_{i=1}^{n} \lambda_i = n\lambda$, n being 365, the number of days per year, and we wait until the first matching birthday occurs. This particular birthday problem is analogous to a ship that goes out to sea with n randomly failing components and <u>no</u> spares; we ask for the probability that a mission is accomplished without loss of function.

The situation described is a classical extreme value problem. Let T_{k_i} represent the random time, measured

from (a) initiation of observation (birthday version), or (b) beginning of mission (logistics version) until the listing of (a) $k_i + 1$ birthdays on day i, or (b) the failure of the last spare for failure-prone element i (we are assuming that the spares are in a cold standby condition; see Gnedenko [12.4]). Then clearly the density function of T_{k_i} is gamma:

$$f_i(x) = e^{-\lambda_i x} \frac{(\lambda_i x)^k}{k_i!} \lambda_i , \quad x \geq 0$$

If T represents the time until n-element system failure (failure = event that at least one element fails with no spare backup), then

$$T = \min_i T_{k_i}$$

and

$$P\{T>t\} = \prod_{i=1}^{n} \left[1 - \int_0^t f_i(x)dx \right] \tag{12.1}$$

In order to simplify (12.1), consider the special case in which $\lambda_i = \lambda$, and $k_i = k$. Put $n\lambda = \mu$, a fixed constant, and study (12.1) as n becomes (realistically) large. Let us scale by n^α and attempt to determine α in such a way that a nondegenerate limiting distribution for T results. Now (12.1) gives

$$P\{T \geq n^\alpha t\} = \left[1 - \int_0^{n^\alpha t} e^{-\lambda x} \frac{(\lambda x)^k}{k!} \lambda dx \right]^n$$

$$= \left[1 - \int_0^{\mu n^{\alpha-1} t} e^{-z} \frac{z^k}{k!} dz \right]^n = (1-I_n)^n \tag{12.2}$$

It now may be shown that if α is selected properly $nI_n \to$ constant for every fixed t and k. In order that nI_n tend to a constant, I_n must tend to zero so $0 < \alpha < 1$ and hence $n^{\alpha-1}t$ is less than unity for sufficiently large n. Therefore, for such an n value,

$$0 < \int_0^{\mu t n^{\alpha-1}} (1-z)\,\frac{z^k}{k!}\,dz < \int_0^{\mu t n^{\alpha-1}} e^{-z}\,\frac{z^k}{k!}\,dz$$

$$< \int_0^{\mu t n^{\alpha-1}} \frac{z^k}{k!}\,dz \qquad (12.3)$$

so

$$\frac{(\mu n^{\alpha-1}t)^{k+1}}{(k+1)!} - \frac{(\mu n^{\alpha-1}t)^{k+2}}{(k+2)k!} < I_n < \frac{(\mu n^{\alpha-1}t)^{k+1}}{(k+1)!} \qquad (12.4)$$

and $nI_n \to$ constant if as $n \to \infty$

$$n \cdot (n^{\alpha-1})^{k+1} \to 1 \qquad (12.5)$$

for which $1 + (\alpha-1)(k+1) = 0$ is required, or

$$\alpha = \frac{k}{k+1} \qquad (12.6)$$

This enables us to conclude that $nI_n \to \frac{(\mu t)^{k+1}}{(k+1)!}$ and

$$P\{T>n^{\alpha}t\} \to e^{-\frac{(\mu t)^{k+1}}{(k+1)!}} \quad \text{for } t \geq 0$$
$$\to 0 \qquad\qquad\qquad\quad t < 0 \qquad (12.7)$$

This leads to an approximate expression for survival of

a mission of length τ ,

$$P\{t>\tau\} \sim e^{-\frac{(\mu\tau)^{k+1}}{(k+1)!}n^{-k}} = e^{-\frac{(\lambda\tau)^{k+1}}{(k+1)!}n} \tag{12.8}$$

which is recognized as a form of the Weibull distribution. Consequently the expected time to failure of the system of n elements is

$$E[T] \sim \int_0^\infty e^{-\frac{(\mu\tau)^{k+1}}{(k+1)!}n^{-k}} d\tau$$

$$= \frac{1}{\mu} n^{\frac{k}{k+1}} [(k+1)!]^{\frac{1}{k+1}} \int_0^\infty e^{-v} v^{-\frac{k}{k+1}} dv \tag{12.9}$$

and

$$\text{Median } T \sim \frac{1}{\mu} n^{\frac{k}{k+1}} [(k+1)! \ \ell n\ 2]^{\frac{1}{k+1}}$$

$$= \frac{1}{\lambda n^{\frac{1}{k+1}}} [(k+1)! \ \ell n\ 2]^{\frac{1}{k+1}} \tag{12.10}$$

Connection with the birthday problem is made by putting $\mu = 1$, $n = 365$, and $k = 1$ in (12.10). We find that the median of the time until first match is about 22.5--close to 23--and so our dynamic birthday problem yields very nearly the same answer as does the classical formulation. The median time to the first occurrence of a <u>triple</u> birthday is predicted to be 82 time units. Relating this to the logistics situation is direct: if a system's elements and spares have mean time to failure of 365 days, say, and 365 different elements are furnished with $k - 1$ spares each then the <u>median</u> length of a <u>system</u> failure-free mission is, according to our approximation, 23 days if $k = 1$, and 82 days if $k = 2$.

Examination of (12.8) shows that if expected failures

over a mission time, $\lambda\tau$, is less than unity then the exponent drops precipitously with increases in k , while if $\lambda\tau$ exceeds unity the exponent may actually increase with k for a time. The source of this anomaly is our initial assumption that $n\lambda = \mu$, a constant, implying that λ must be small in order that the approximation be valid. We should therefore beware of (12.8) when $\lambda\tau > 1$, and trust its accuracy especially when $\lambda\tau << 1$. Fortunately, this case is of considerable interest in many spares stocking problems of logistics.

Expression (12.8) suggests that the optimal stocking of noninterchangeable, exponentially failing, spares can be expressed as a nonlinear integer programming problem: consider g groups of elements, each of which has characteristic failure rate λ_i $(i=1,2,\ldots,g)$. The number of elements in group i is n_i . To stock, choose k_i $(i=1,2,\ldots,g)$ to solve the problem

$$\min_{k_i} \sum_{i=1}^{g} \frac{\lambda_i^{k_i+1}}{(k_i+1)!}\, n_i$$

subject to, say

$$\sum_{i=1}^{g} n_i(k_i-1) \leq K$$

$$\sum_{i=1}^{g} n_i(k_i-1)w_i \leq W \quad \text{(no more than } W \text{ lbs, or dollars, invested)}$$

One can also confine solutions to the region (in $k_1,k_2,\ldots,k_g$ space) for which

$$\frac{\lambda_i^{k_i+1}}{(k_i+1)!}\, n_i \leq U_i \qquad (12.11)$$

where U_i defines a <u>chance constraint</u> for the ith

group: if (12.11) is satisfied then the probability that any member of that group fails during a mission (of unit length) does not exceed e^{-U_i}. We do not offer algorithms for solution of these mathematical programming problems at present.

Of course, alternative formulations of the above setup suggest themselves in abundance. Some randomly selected targets of opportunity may be quickly disposed of. For instance, suppose that failures are not exponential but are gamma with shape parameter one, so that $f(x) = e^{-(\lambda/2)x}(\lambda/2)^2 x$, yielding mean time to failure of λ^{-1} once again, but displaying a wear-out characteristic. Then norm by setting $n(\lambda/2) = \mu$ and replace k by $2k$ in our various expressions (12.2-.10) to obtain a limiting form like (12.8). Notice that since behavior of the density f near $x = 0$ inevitably governs the minimum, it is not necessary to insist on the form of f throughout; this is familiar in the theory of extreme values. The assumption that spares fail, that is, that standbys are "hot," yields

$$P\{T>t\} = \left[1 - \left[\int_0^t f(x)\,dx\right]^k\right]^n$$

$$= [1 - (1-e^{-\lambda t})^k]^n$$

$$\sim e^{-nk\lambda t}$$

if $n\lambda = \mu$ constant as $n \to \infty$.

I point out that the preceding models simply permitted the replacement of failing elements, and did not consider the possibility of repair. Under present circumstances repair personnel are an expensive resource, and it will be of interest to examine strategies that range between labor intensive (ample repair personnel) and materiel intensive (mainly replacement) extremes. Models involving repair--that are to a greater degree labor intensive--will be considered subsequently.

12.3 An Adaptive Decision Problem Involving a Markov Chain

Again, let us consider the situation in which failure-prone elements must be backed up by spares, but in addition suppose (i) that failure rates differ between elements, may actually change occasionally, and are unknown, and hence (ii) a fixed spares allocation cannot be relied upon forever. We propose and analyze a very simple but adaptive spares allocation process that we call "play the loser." Our illustration is too simple to be immediately useful, but may perhaps be elaborated upon with profit. No doubt it is but a country cousin of the Markov decision processes of Howard [12.5].

Begin by assuming that exactly two elements are of concern, and a spare for only one of these may be budgeted. Under Configuration 1 the spare is allotted to Element 1, while under Configuration 2 the spare backs up Element 2. Suppose that time is measured in terms of missions (of length τ, say), and that the spare assignment may be made anew at each mission's beginning.

The general idea is as follows. Imagine that Configuration 1 is in effect (spare allocated to Element 1, no spare to Element 2) at the beginning of a mission. Let us examine the possible events that may occur during the mission, and ask ourselves which of these might suggest the wisdom of changing to Configuration 2 for the next mission. The possibilities are shown in Table 12.1 where we use the symbol "F" to denote a failure, and "$\not$F" to denote no failure. Offhand, there is no

Table 12.1 Possible Events during the Mission

Case	Component 1	Component 2
1.	$\not$F	$\not$F
2.	$\not$F	F
3.	F $\not$F	$\not$F
4.	F $\not$F	F
5.	F F	$\not$F
6.	F F	F

reason to make a change unless Element 2 fails; if this event occurs in concert with at most one failure of Element 1 there is reason to consider Element 2 the more failure prone ("the loser") and hence we adopt the following:

Decision Rule. If the spare is assigned to Element 1 on a mission and if Cases 2 or 4 occur, then switch the spare to Element 2 on the next mission. Otherwise continue in Configuration 1. The same rule, with the names of the elements reversed, prevails if the spare is initially assigned to Element 2.

We shall work out the implications of this rule for the case of constant failure rates, after pointing out that it is obviously quite adaptive. If, say, the failure rate of Element 2 changes for the worse, our rule will automatically allocate the spare to Element 2 with higher probability. Of course, an occasional freak outcome will send the spare back to Element 1, but this will happen infrequently. Various refinements of this rule are possible, such as the use of the last two missions' history, and the addition of cost or essentiality indices of the elements.

Analysis of the One-Stage Play-the-Loser Rule. Because of our assumption that failures occur independently and exponentially the imposition of the rule means that the spare location is a simple Markov chain with stationary transition probabilities. Let $U_1(t)$ denote the probability that Configuration 1 is in force at the beginning of patrol or mission t, and $U_2(t)$ be the probability of Configuration 2. The one-step transition probability matrix clearly appears as shown in Table 12.2. The usual conditional probability arguments relating to Markov chains now show that

$$U_1(t+1) = U_1(t)p_{11} + U_2(t)p_{21} \qquad (12.12)$$

Since the chain is irreducible and ergodic there is a stationary or long-run distribution denoted by U_i $(i= 1,2)$,

Table 12.2 One-Step Transition Probability Matrix

at t	Configuration at t + 1: 1	2
1	$1 - p_{12} = p_{11}$	$p_{12} = \dfrac{(1+\lambda_1 M)(1-e^{-\lambda_2 M})}{e^{\lambda_1 M}}$
2	$p_{21} = \dfrac{(1+\lambda_2 M)(1-e^{-\lambda_1 M})}{e^{\lambda_2 M}}$	$1 - p_{21} = p_{22}$

$$U_i = \lim_{t\to\infty} U_i(t)$$

Since $U_2 = 1 - U_1$ we get immediately from (12.12) the information that

$$U_1 = \frac{p_{21}}{p_{21} + p_{12}}$$

Now the system reliability, given Configuration 1, is

$$R_1 = e^{-\lambda_1 M}(1+\lambda_1 M)e^{-\lambda_2 M}$$

while given Configuration 2 it is

$$R_2 = e^{-\lambda_2 M}(1+\lambda_2 M)e^{-\lambda_1 M}$$

Hence the long-run system reliability obtained by following our one-stage play-the-loser rule is

$$R = U_1T_1 + U_2T_2$$

Numerical studies of the operation of the decision rule have been made, and indicate that it is quite effective. Of course, a realistic application would involve changes across many candidates for spares. Other realistic touches will undoubtedly occur to the reader, but we feel that this example illustrates a somewhat novel application of Markov chain methods in the context of logistics.

12.4 Manpower and Personnel-Related Problems: Two Simple Examples

Owing to increases in the expense of recruiting and training high quality support personnel, there is considerable reason for the Navy, and the Department of Defense in general, to manage the assignments of such personnel as wisely as possible. It seem obvious, for example, that the presence of adequate skilled maintenance (and operating) manpower aboard ships, or at shore facilities, will increase the availability of complex new equipments and hence force effectiveness. On the other hand, policy dictates that assignments not be indefinitely long: an incumbent must be allowed to rotate from ship to shore and back after a reasonable time period (usually several years). It can be anticipated that many a new assignee will not have the training and experience necessary to cope adequately with his tasks, and that a period of learning must be anticipated. Our model attempts to highlight the tradeoff between advantages of rather long tours of duty, during which on-the-job training and working experience may increase effectiveness, and the necessities of rotation.

Model 1. Rotation Frequency and its Effect upon Availability. Suppose that when a particular piece of equipment (for example, sonar, communications gear) is in use it is prone to random failure at rate λ, and then to repair at rate μ. Further, λ and μ values are influenced by experience of the repairman: If a man is inexperienced his preventive maintenance capabilities are low, so $\lambda = \lambda_1$, and his repair rate is also relatively low, that is, $\mu = \mu_1$. After a time S

he gains experience and $\lambda = \lambda_2 < \lambda_1$, while $\mu = \mu_2 > \mu_1$.

Now assume that it is a policy to rotate or perhaps replace the repairman every R units of time, for example, when enlistment is up, where R may be of order 3-4 years. Assume also that when a new repairman comes aboard, he is inexperienced. Our objective is to study the dependence of the system availability--the probability that the system is up--upon R. Clearly, lengthening R may be expensive, but it should provide higher system availability, and we investigate the tradeoff.

The model examined is simplified. Features that, among others, influence availability are (i) quality of personnel, as reflected, for instance by μ_1 and μ_2 values and the time to change from the low rate to a higher one, (ii) availability of shore or tender repair facilities, (iii) the number of repair personnel assigned to carry out repair, and (iv) the design of the equipment, since the latter influences failure rate and repair rate, that is, the workload. The following steps lead to formulas for studying availability versus R.

(1) Let "1" denote the Up state, and "0" the Down state. Then

$$a_{11}(t) = \text{Prob}\{\text{system Up (in State 1) at } t \mid \text{Up at 0}\}$$

$$a_{10}(t) = \text{Prob}\{\text{system Down (in State 0) at } t \mid \text{Up at 0}\}$$

the a-transition probabilities apply to the <u>inexperienced</u> man; we find

$$a_{11}(t) = \frac{\mu_1}{\lambda_1+\mu_1} + \frac{\lambda_1}{\lambda_1+\mu_1} e^{-(\lambda_1+\mu_1)t}$$

$$a_{10}(t) = \frac{\lambda_1}{\lambda_1+\mu_1}\left[1-e^{-(\lambda_1+\mu_1)t}\right]$$

$$a_{01}(t) = \frac{\mu_1}{\lambda_1+\mu_1}\left[1-e^{-(\lambda_1+\mu_1)t}\right]$$

$$a_{00}(t) = \frac{\lambda_1}{\lambda_1+\mu_1}\ \frac{\mu_1}{\lambda_1+\mu_1}\ e^{-(\lambda_1+\mu_1)t}$$

(2) The corresponding probabilities applying to the experienced man are as follows.

$$b_{11}(t) = \frac{\mu_2}{\lambda_2+\mu_2} + \frac{\lambda_2}{\lambda_2+\mu_2}\ e^{-(\lambda_2+\mu_2)t}$$

$$b_{10}(t) = \frac{\lambda_2}{\lambda_2+\mu_2}\left[1-e^{-(\lambda_2+\mu_2)t}\right]$$

$$b_{01}(t) = \frac{\mu_2}{\lambda_2+\mu_2}\left[1-e^{-(\lambda_2+\mu_2)t}\right]$$

$$b_{00}(t) = \frac{\lambda_2}{\lambda_2+\mu_2} + \frac{\mu_2}{\lambda_2+\mu_2}\ e^{-(\lambda_2+\mu_2)t}$$

(3) In our initial model, let us start a cycle (of length R) with an inexperienced man, and then calculate

$$P_{11}(R) = \text{Prob}\{\text{system up at } R \mid \text{system up at } 0\}$$

$$= P\{U(R) = 1 \mid U(0) = 1\}$$

We have let

$U(t) = 1$, if system up at t

$= 0$, if system down at t

Let us denote by S the time at which "experience sets in," that is, when rates change favorably. The formula for $P_{11}(R)$ is as follows

$$P_{11}(R) = a_{11}(S)b_{11}(R-S) + a_{10}(S)b_{01}(R-S) \qquad (12.13)$$

if $S < R$, in which case there is time to attain the level of experience yielding the lower failure rate λ_2, and higher repair rate λ_1. The argument is that if the system is up initially it may (i) be up at S, when "experience sets in," and in turn be up at R, having been up at the change point. Notice that we are utilizing the Markov property of the process in writing (12.13). Alternatively, (ii) the system may be up initially, be down at S, and then be up at R. We can then evaluate $P_{11}(R)$ explicitly in terms of our formulas for the a's and b's; $P_{10}(R) = 1-P_{11}(R)$. Note that if $R < S$, so rotation occurs before skill level increases, then

$$P_{11}(R) = a_{11}(R) \qquad (12.14)$$

Similarly,

$$P_{01}(R) = a_{01}(S)b_{11}(R-S) + a_{00}(S)b_{01}(R-S) \qquad (12.15)$$

and

$$P_{00}(R) = 1 - P_{01}(R)$$

Fancier models, in which S is a random variable, can be worked out with only a little trouble. So can situations in which μ_1 and μ_2, λ_1 and λ_2 are drawn from some distribution.

(4) Probably S and R are of the order of years (for example, $S \gtrsim 1$ year, $R \gtrsim 2$ or more). However, for many equipments the mean time between failures (MTBF) λ^{-1} may be of the order of days or weeks, the same being true for μ^{-1} (the latter may be less than one day). This fact suggests that we can sometimes use the steady state values (disregard exponential terms in expressions for a's and b's). This will simplify results.

Now let us derive consequences. Consider first the short run. Suppose the equipment is up initially (just been serviced by an expert). Let us look at average availability over one R-cycle: actual availability at t is $U(t)$--one or zero at t--so expected average availability is

$$\alpha_R = E\left\{\frac{1}{R}\int_0^R U(t)dt\right\} = \frac{1}{R}\int_0^R E[U(t)]dt$$

$$= \frac{1}{R}\left\{\int_0^S a_{11}(t)dt + a_{11}(S)\int_0^{R-S} g_{11}(t)dt\right.$$

$$\left. + a_{10}(S)\int_0^{R-S} b_{01}(t)dt\right. \qquad (12.16)$$

In view of comments (4), drop the exponentials and integrate.

$$\alpha_R \simeq \frac{1}{R}\left\{\int_0^S \frac{\mu_1}{\lambda_1+\mu_1}dt + \frac{\mu_1}{\lambda_1+\mu_1}\int_0^{R-S}\frac{\mu_2}{\lambda_2+\mu_2}dt\right.$$

$$\left. + \frac{\lambda_1}{\lambda_1+\mu_1}\int_0^{R-S}\frac{\mu_2}{\lambda_2+\mu_2}dt\right.$$

$$= \frac{S\left(\frac{\mu_1}{\lambda_1+\mu_1}\right) + (R-S)\left(\frac{\mu_2}{\lambda_2+\mu_2}\right)}{R}$$

$$= \frac{S}{R}\left(\frac{\mu_1}{\lambda_1+\mu_1} - \frac{\mu_2}{\lambda_2+\mu_2}\right) + \frac{\mu_2}{\lambda_2+\mu_2} \qquad (12.17)$$

Numerical example. Suppose

$\frac{\mu_1}{\lambda_1+\mu_1}$ = "inexperienced availability" = 0.65

$\frac{\mu_2}{\lambda_2+\mu_2}$ = "experienced availability" = 0.95

If $\frac{S}{R} = \frac{1}{2}$,

$\alpha_R \simeq \frac{1}{2}(0.65-0.95) + 0.95 = 0.80$

If $\frac{S}{R} \simeq \frac{1}{5}$,

$\alpha_R = 0.89$

If $\frac{S}{R} = \frac{1}{10}$,

$\alpha_R \simeq 0.92$

Clearly, availability is penalized by rapid turnover of personnel.

In the long run, that is, after several R-cycles (repairman replacements) have occurred, the probability that the system is up initially is not one, as has just been assumed. The state of the system at moments of personnel change is actually a two-state Markov chain,

with (12.13), (12.14), and (12.15) supplying the one-step transition probabilities.

Let P_1 be the long-run or steady state probability that the system is up at the beginning of an R cycle. Then P_1 satisfies an equation of probability balance: put $Q_1 = 1 - P_1$, then

$$P_1 \cdot P_{10}(R) = Q_1 P_{01}(R)$$

$$= (1-P_1)P_{01}(R)$$

so

$$P_1 = \frac{P_{01}(R)}{P_{10}(R) + P_{01}(R)} \tag{12.18}$$

Now one can evaluate this probability by means of (12.13-.15). Then the job is to evaluate a <u>modification</u> of (12.16) that accounts for the fact that the system may be in one of two states. If $\alpha_R(\infty)$ denotes the long-run availability,

$$\alpha_R(\infty) = \frac{1}{4}\Bigg\{\int_0^S [P_1 a_{11}(t) + Q_1 a_{01}(t)]dt$$

$$+ [P_1 a_{11}(S) + Q_1 a_{01}(S)] \int_0^{R-S} b_{11}(t)dt$$

$$+ [P_1 a_{10}(S) + Q_1 a_{00}(S)] \int_0^{R-S} b_{01}(t)dt\Bigg\} \tag{12.19}$$

The tedious evaluation requiring exponentials is probably not required in view of comments under (4) above about magnitudes of R , S , λ^{-1} , and μ^{-1} .

We can derive useful approximate formulas. If $(\lambda+\mu)S$ and $(\lambda+\mu)(R-S)$ are large (safely, at least 10), then

$$P_{01}(R) \simeq \frac{\mu_1}{\lambda_1+\mu_1} \cdot \frac{\mu_2}{\lambda_2+\mu_2} + \frac{\lambda_1}{\lambda_1+\mu_1} \frac{\mu_2}{\lambda_2+\mu_2} = \frac{\mu_2}{\lambda_2+\mu_2}$$

and substitution into (12.18) gives

$$P_1 \simeq \frac{\mu_2}{\lambda_2+\mu_2} \quad , \quad Q_1 \simeq \frac{\lambda_2}{\lambda_2+\mu_2}$$

Then if one substitutes into (12.19) ignoring all exponential terms the result once again is (12.17):

$$\alpha_R(\infty) \simeq \frac{S}{R}\left(\frac{\mu_1}{\lambda_1+\mu_1} - \frac{\mu_2}{\lambda_2+\mu_2}\right) + \frac{\mu_2}{\lambda_2+\mu_2}$$

Thus, effectively the "long run" is achieved nearly immediately for the situations we consider here.

The model and analysis contain several implicit assumptions that are worth review. These can be relaxed if need be.

(a) Equipment is assumed to fail and be repaired "at random," that is, with an exponential distribution. One can ask whether actual times between failure are influenced by equipment usage, for an equipment is not likely to fail if it is "off," that is, not in use. Further models that include consideration of usage patterns can be constructed if desired. Actual supporting data will be of use in giving truly credible support to the models; perhaps the Navy's Maintenance and Material Management (3M) system can furnish such data. However, the models can also use subjective or judgmental inputs. If desired, subjectivity of input can be quantified by means of the Bayesian methods of statistics. With respect to availability, our present formulas apply even when quite general time to failure and repair time distribution prevail. To a good first approximation, the exponential assumptions are of secondary importance.

(b) Our present models assume that repair action begins as soon as failure is detected. That is, there is no delay because of (i) the unavailability of spare

parts, or (ii) the unavailability of expert supervisory talent. One possible way of handling the effect of skilled supervision is solely data-analytic and pragmatic: we obtain data on ships or installations that utilize the same equipments but that have different complements of repair talent. Then we compute estimates of μ and λ under the differing conditions, and compare. Unfortunately, such data are not readily available, and, if they are, may not lead to conclusive comparisons because of the influence of other factors. Nevertheless, the analysis of such data will give a useful check on conclusions derived from analytical models, or such methods as the SHIP II simulation developed by the Naval Personnel Research and Development Laboratory (now Center) of San Diego, California.

Model 2. Doctors on Ships? Our next model is developed in response to the question: Should expensive specialists, for example, medical doctors, be assigned to ships beginning a mission? The issues that arise are as follows.

(1) If a man is injured or becomes seriously ill while a mission is in progress, and if no doctor is present, it may be necessary to transfer the individual from the ship to a hospital for treatment. Presence of a doctor will, at least in some cases, permit the treatment to take place on the ship. Thus, the cost of the transfer may be eliminated--at, of course, the expense of maintaining the doctor. Similarly, but not entirely analogously, breakdown of key equipment entails loss of military effectiveness and may require that the mission be prematurely terminated. If a skilled specialist is aboard such events may be forestalled.

(2) The cost of retaining doctors or other trained specialists is very high. The cost for such a specialist may rightfully include some of the expense of his original recruitment and training.

The Occurrence of Demands. Basic to the question of whether a doctor or highly trained and expensive repairman should be added to a ship's complement is the likelihood of demand for services that he alone can supply. We propose some probability models for this question. Actually, a complex variety of sources may conspire to cause demand.

Model 2-A. Simple Chance Demand, Caused by Accidents or Sudden Disease. Imagine that a ship has n individuals aboard when it sets out on a mission of duration M. Each individual is thought to have a constant probability λdt of experiencing an accident or sudden severe illness between t and $t + dt$, dt being a small number. The occurrence of accidents or illness is first assumed to be independent from individual to individual. Then our simple model implies that each individual experiences his demanding event at time T_i $(i=1,2,\ldots,n)$ measured from the start of the mission, T_i being distributed in accordance with an exponential distribution with rate parameter λ:

$$P\{T_i \leq t\} = 1 - e^{-\lambda t} \quad , \quad 0 \leq t < \infty$$
$$= 0 \quad , \quad t < 0$$

Next, the occurrence of the smallest T_i on the ship of crew size n is the distribution of the minimum of a sample of n independent T_i's, which is exponential with parameter n. Let us call this time τ_n; then

$$P\{\tau_n > t\} = e^{-n\lambda t} \quad , \quad 0 \leq t < \infty$$
$$= 1 \quad , \quad t < 0$$

According to this model, if no doctor (or repairman) is present:

(A) The probability that the mission does not terminate during M for the cause associated with λ is

$$P\{\tau_n > M\} = e^{-n\lambda M}$$

(B) The expected time to the scheduled end of a mission that involves a rescue (or terminates early) is

$$E[M-\tau_n \mid \tau_n<M] = \frac{\int_0^M (M-x)e^{-n\lambda x}\, n\lambda dx}{1 - e^{-n\lambda M}}$$

$$= \frac{M}{1 - e^{-n\lambda M}} - \frac{1}{n\lambda} \qquad (12.20)$$

It is of interest to see what this formula approaches as λ becomes small, a condition likely to be true in practice. Write (12.20) as

$$E[M-\tau_n \mid \tau_n<M] = \frac{e^{-n\lambda M} - 1 + n\lambda M}{(n\lambda)(1-e^{-n\lambda M})} \qquad (12.21)$$

and then expand in Taylor's series to obtain

$$P[M-\tau_n \mid \tau_n<M] = \frac{\frac{(n\lambda M)^2}{2!} + o(n\lambda M)}{n\lambda(n\lambda M+o(n\lambda M))}$$

$$\rightarrow \frac{M}{2}$$

This states that in the limit of vanishingly small accident or disease rate an average of one-half the mission time will be lost, provided that an event occurs ($\tau_n<M$) and that the probability is very small of an accident or breakdown of the sort envisioned (for example, a heart attack, stroke, severe injury, in the medical case).

Model 2-B. Chance Demand, Differing Demand Rates. We can realistically generalize Model 2-A to the situation for which each individual has a different characteristic rate or probability of requiring the vital service: $\lambda_j dt$ is essentially the demand probability for individual j ($j=1,2,\ldots,n$) . Then the overall demand rate is, assuming independence, equal to

$$\lambda(n) = \lambda_1 + \lambda_2 + \cdots \lambda_n$$

and the distribution of τ_n is still exponential, with $\lambda(n)$ replacing in (12.20). All this is well known; see Feller [12.1].

Practically speaking, one might consider classes of individuals who are more or less susceptible to disease or accident, and who can be characterized by the same failure or catastrophe rates within the class. For example, the older officers (ship captain, executive officer, and so on) may be less likely to incur an appendicitis outbreak than would a younger man; on the other hand, an older person might be more susceptible to heart attack. This sort of consideration would suggest putting individuals into classes, with characteristic rates $\{\lambda'_u, u = 1,2,\ldots,c\}$; then

$$\lambda(n) = \sum_{u=1}^{c} n(u)\lambda'_u$$

where $n(u)$ is the number of individuals in the uth class.

Medical data might well be available to provide estimates for the above rates. For equipments, 3M data might well be interpreted for the same purpose.

$$P\{\tau_n > x | \lambda_1, \lambda_2, \ldots, \lambda_n\} = e^{-\left(\sum_{j=1}^{n} \lambda_j\right)x}$$

Now, by assumption, each λ_j is independently drawn from a population with distribution $H(y)$, and density $h(y)$. Consequently, removal of the condition on the λ's amounts to integrating:

$$P\{\tau_n > x\} = \prod_{j=1}^{n} \int_0^{\infty} e^{-yx} h(y)dy = [\hat{h}(x)]^n$$

where $\hat{h}(x)$ represents the Laplace transform of the density h.

Example. Let h be a Gamma density

$$h(y) = \frac{e^{-\alpha y}(\alpha y)^{\beta-1}}{\Gamma(\beta)}$$

Then

$$\hat{h}(x) = \left(\frac{\alpha}{\alpha+x}\right)^{\beta}$$

and the chance that a mission proceeds with no demand is

$$P\{\tau_n > M\} = \left(\frac{\alpha}{\alpha+M}\right)^{\beta n}$$

It will also be interesting to derive the expected lost mission time under this model. It is

$$E[(M-\tau_n)^+ | \lambda_1, \lambda_2, \ldots, \lambda_n] = \int_0^M (M-x) e^{-\lambda(n)x} \lambda(n) dx$$

Next, the conditions on the λ's must be removed. This can be done easily in terms of our gamma density illustration. The result is

$$E[(M-\tau_n)^+] = M - \frac{\alpha}{\beta n - 1} + \frac{1}{\beta n - 1}(\alpha+M)\left[\left(\frac{\alpha}{\alpha+M}\right)\right]^{\beta n}$$

A generalization can be carried out for the case in which several subgroups of crew members are described by their specific gamma distributions.

Armed with models that describe demand for service by a medical doctor or other specialist we can formulate decision analyses. Our demand models provide inputs to

these analyses, as do certain costs.

Decision Model 1. Suppose ship missions are of approximately constant duration M. Let D be the (dollar) cost per unit time of maintaining a medical doctor aboard ship. Then MD is the dollar cost per mission of keeping the doctor aboard ship while the ship is engaged in an active mission.

Let R be the cost of the evacuation or rescue operation necessary when an emergency arises and no doctor is present. Think of R as being an average cost; clearly this cost will vary with the location of the ship, and also with the individual, that is, the recipient of the rescue.

Let $p(n;M)$ denote the expected number of emergencies that arise when a crew of n individuals embarks on a mission of duration M. Our previous models provide various bases for evaluating $p(n;M)$.

Model 2-A implies that the number of demands during a mission is binomially distributed with mean $p(n;M) = n(1-e^{-\lambda M})$. Suppose that each emergency requires a separate rescue or evacuation operation. Then the expected cost of rescues or evacuations is $p(n;M)R$ per mission if a doctor is not aboard. Presume that if a doctor is aboard all of these can be avoided, but at cost MD. The optimal decision rule is then

Carry a doctor if $MD < n(1-e^{-\lambda M})R$

Do not carry a doctor if $MD > n(1-e^{-\lambda M})R$

If λ is quite small, as should often be true, this becomes a good approximation:

Carry a doctor if $D < n\lambda R$

Do not carry a doctor if $D > n\lambda R$

Of course if there is equality ($D = n\lambda R$, for example) then other considerations must settle the matter.

Decision Model 2. This model simply recognizes the differences between demand (injury, accident, or

sickness) rates between individuals, as in Demand Model 2-B. For that model the expected number of demands is

$$p(n;M) = \sum_{j=1}^{n} (1-e^{-\lambda_j M})$$

Hence our decision rule becomes

Carry a doctor if $MD < R \sum_{j=1}^{n} (1-e^{-\lambda_j J})$

Do not carry a doctor if $MD > R \sum_{j=1}^{n} (1-e^{-\lambda_j M})$

Some additional comments may be made on these models.

(A) The models tacitly assume that R, the cost of a rescue operation, is the same regardless of mission. In fact, one can assign a cost that depends upon the mission and then decide on the basis of our various decision models whether a doctor can be justified.

(B) The same comment as in (A) above holds for the rates or λ-values likely to prevail on different missions.

(C) The above decision rules, derived for ships, can apply also to groups of ships. The doctor can be located on one ship of the group, and emergencies will then be transferred to that ship when they occur.

(D) In the above discussion the λ-values are taken to be known. To make the decision we must obtain estimates, and then treat these estimates as equal to the parameter values actually prevailing. A more sophisticated approach explicitly recognizes that estimates are uncertain; one standard way of handling that situation is by means of Bayesian decision theory.

The above decision models assume that emergencies generate rescue costs, but do not shorten missions. In other situations, perhaps having to do with the failure of a major weapon system, this might not be the case. It may well be that if a major system goes out on, say, a submarine, the latter must return to port prematurely. We set up a simple and tentative model for this situation, anticipating that refinements in the model may suggest themselves.

Decision Model 3. Suppose the initial cost for a copy of the ship in question is S (dollars), and that the anticipated life is equivalent to L missions of length M. It is reasonable to assess a penalty of S/LM dollars per unit time that the ship is not carrying out its assigned task during a mission, owing to lack of specialized repair personnel or spare parts.

The expected lost time per mission of length M is obtained for Demand Model 2-A by multiplying (12.21) by the probability of at least one demand during M, namely $1 - e^{-n\lambda M}$. Thus,

$$E[\max(M-\tau_n,0)] = M - \left[\frac{1-e^{-n\lambda M}}{n\lambda}\right]$$

Thus, the expected cost of lost service if a specialist, or requisite spares, are not carried over the life of the ship is

$$\text{Expected cost} = \left(\frac{S}{LM}\right)L \cdot E[\max(M-\tau_n,0)]$$

$$= S\left\{1 - \left[\frac{1 - e^{-n\lambda M}}{n\lambda M}\right]\right\}$$

The optimal decision rule is then derived from the principle that one should carry the specialist if his total cost over the life of the ship MLD is less than the expected cost of curtailed missions:

Carry specialist if $$MLD < S\left\{1 - \left[\frac{1 - e^{-n\lambda M}}{n\lambda M}\right]\right\} \quad (12.22)$$

Do not carry specialist if $$MLD > S\left\{1 - \left[\frac{1 - e^{-n\lambda M}}{n\lambda M}\right]\right\}$$

A very similar formula can be written down if Demand Model 2-B is invoked. Note that n now refers to the number of failure-prone elements to be serviced by the specialist.

One qualitative fact that emerges from (12.22) is that for fixed total mission time $ML = T$ one can reduce the need for a specialist by shortening mission time, M, (and correspondingly increasing L). By indefinitely shortening M the right-hand side of (12.22) can be brought very close to zero, which guarantees that our decision rule will recommend that the specialist be left ashore. Of course, indefinite shortening of the mission time is impractical, but the tendency is of interest and can be quantitatively assessed by use of formulas like (12.22).

Elaborations of the above models may be made to reflect (a) lack of independence between events of disease, injury, or equipment failure, for example, as caused by the outbreak of an epidemic or a fire or occurrence of a military engagement.

12.5 Repair Models

The situations that are considered in this section are logical extensions and elaborations of some of those previously discussed. Our concern is with the availability status of a unit (ship, airplane, submarine, satellite, and so on) that carries a multitude of different but interrelated subsystems, together with support personnel and backup spares. When a unit first departs on a mission it is likely to be in a better condition of readiness or system availability than is the case as the mission progresses, owing to superior facilities at base or tender. Such facilities include considerable diagnostic and repair talent and equipment, as well as spares. The models that we introduce are

aimed at representing the degradation of overall ship system performance as the mission progresses. It will be interesting to relate this degradation to the quality and numbers of personnel aboard ship (including the quality of leadership, and organizational structures), to equipment design (extent of modularization, for example), and to the frequency of overhauls or tender visits. Empirical investigation of these questions seems possible to initiate by making use of the 3M data base, although difficulties arise. This section considers certain mathematical models that may be a useful framework for conceptual and data analysis and prediction. They are oversimplified but flexible. An alternative analytical tool is the so-called SHIP II simulation model developed by the Naval Personnel Research and Development Laboratory (now Center). This simulation model attempts to represent many of the real-life constraints and priorities that affect shipboard maintenance, but is quite complex and requires extensive computer running time. Our analytical models provide useful supplementary information at a much reduced cost of manipulation.

Model 1. Suppose that a repair force is assigned to handle failures that occur in a particular section of the ship. Several repair forces may be present--one for each of several sections aboard ship. When failures occur, the crew immediately begins service. In the present model we assume that certain equipment failures are secondary or non-disabling, that is, merely degrade equipment performance, while others are killing or disabling, bringing the equipment down immediately. At such moments the repair force immediately begins equipment repair; repair action means that both the killing and the secondary failures are repaired, returning the equipment to an essentially new condition.

Let $\underline{U}_i$ denote the ith up or available period, and $\underline{R}_i$ the ith repair period (which immediately follows the termination of $\underline{U}_i$). Then it is clear that for our model $(\underline{U}_i, \underline{R}_i;\ i = 1,2,\ldots)$ is a sequence of independent pairs of random variables, but $\underline{U}_i$ and $\underline{R}_i$ have tendency to be positively correlated. Properties

of the $(\underline{U}_i, \underline{R}_i)$ renewal process are derived in the dissertation by Luckew [12.10], and will be summarized here.

(a) $(\underline{U}_i+\underline{R}_i)$ is an ordinary renewal process.

(b) Suppose the killing event occurs exponentially, so $f_{\underline{U}}(x) = \lambda e^{-\lambda x}$ is the density of an up time; $\underline{R}$ is the sum of the repair times of killing and secondary events. Then the Laplace transform of the joint density of an up time and subsequent down time is

$$\hat{f}_{\underline{U},\underline{R}}(s,\zeta) = E[e^{-s\underline{U}}e^{-\zeta\underline{R}}] = \frac{\lambda\hat{\ell}_0(\zeta)[1-\hat{k}(s+\lambda)]}{(s+\lambda)[1-\hat{\ell}(\zeta)\hat{k}(s+\lambda)]} \quad (12.23)$$

where $\hat{\ell}_0$, $\hat{k}$, and $\hat{\ell}$ are respectively the Laplace transform of the densities of the repair time of the killing event, the interarrival times of secondary events, and the repair times of secondary events.

(c) Moments are available from (12.23):

$$E[\underline{U}] = \frac{1}{\lambda} \qquad E[\underline{R}] = \frac{E[R]}{1-\hat{k}(\lambda)}$$

$$Var[\underline{U}] = \frac{1}{\lambda^2} \qquad Var[\underline{R}] = \frac{Var[R]}{1-\hat{k}(\lambda)} + (E[R])^2 \frac{\hat{k}(\lambda)}{[1-\hat{k}(\lambda)]}$$

$$cov[\underline{U},\underline{R}] = -\frac{E[R]\hat{k}'(\lambda)}{[1-\hat{k}(\lambda)]^2}$$

where R is a secondary failure repair time, and $k'(\lambda)$ is the derivative of $\hat{k}$, evaluated at λ; since the latter is negative, the covariance is positive as anticipated.

(d) Consider the renewal process generated by the alternation of $\underline{U}_i + \underline{R}_i$, $i = 1,2,\ldots$. Let $U(t)$ represent the probability that the system is up at t , conditional on just entering the up state at $t = 0$. Then, from renewal theory,

$$u(\infty) = \lim_{t\to\infty} \frac{E[\underline{U}]}{E[\underline{U}+\underline{R}]}$$

Explicit formulas come from (c).

(e) Let V_t be the forward recurrence time until the end of a repair period, and let $r(w,t)$ be its density:

$$r(w,t) = \lim_{\Delta w\to 0} \frac{P\{V_t \varepsilon (n,w+\Delta w), t\varepsilon(\underline{R})\}}{\Delta w}$$

where $t\varepsilon(\underline{R})$ signifies that the system is in the repair state at time t . The Laplace transform of the long-run density may be shown to be (here $\hat{\ell}_0(\zeta) \equiv \hat{\ell}(\zeta)$)

$$\hat{\underline{r}}(\zeta) = \frac{1}{E[\underline{U}+\underline{R}]} \frac{1-\hat{\ell}(\zeta)}{\zeta[1-\hat{\ell}(\zeta)\hat{k}(\lambda)]}$$

from which moments may be derived. Furthermore, let a repair period be in progress at t ; then, in the long run, the number of repairs completed at the instant of observation is geometrically distributed if $\underline{U}$ is exponentially distributed. Similarly, the number of secondary event failures accumulated before an observation during a long-run up time is geometric.

Model 2. In this model we shall again differentiate between secondary failure that may degrade but not disable performance, and a killing failure. However, we shall assume that the secondary failures become apparent immediately--not just when a killing failure occurs. Furthermore, we suppose that secondary failures

are repairable, although they may be required to wait for repair facilities to become available.

Now, if the overall secondary failure process is approximately Poisson with parameter α, and if one crew is furnishing repairs with expected repair time $E[R]$, then the repair queue is $M/G/1$, and as $\rho = \alpha E[R]$ becomes close to, but below, unity, a diffusion approximation for backlogged repair jobs is suitable. Let $\underline{R}(t)$ denote total accumulated repair time at time t after an instant at which no equipment is down, and let $\underline{U}$ denote the time until a killing event occurs; $\underline{U}$ has exponential density $\lambda e^{-\lambda y}$. Now it may be shown, see [12.2] that the total accumulated repair time at $\underline{U}$, $\underline{R}(\underline{U})$, is <u>exponentially distributed</u>:

$$f_{\underline{R}(\underline{U})}(x) = -\eta(\lambda)e^{\eta(\lambda)x}$$

where

$$\eta(\lambda) = \frac{\mu}{\sigma^2}\left[1 + \left(1 + \frac{2\sigma^2\lambda}{\mu^2}\right)^{\frac{1}{2}}\right], \text{ if } \rho < 1$$

$$= \frac{\mu}{\sigma^2}\left[1 - \left(1 + \frac{2\sigma^2\lambda}{\mu^2}\right)^{\frac{1}{2}}\right], \text{ if } \rho > 1$$

Here μ and σ^2 are the infinitesimal mean and variance of the approximate diffusion

$$E[\underline{R}(t+h)-\underline{R}(t)] = \mu h + 0(h) \approx (\alpha E[R]-1)h \equiv (\rho-1)h$$

$$\text{Var}[\underline{R}(t+h)-\underline{R}(t)] = \sigma^2 h + 0(h) \approx \alpha E[R^2]h$$

By means of these results one can estimate the backlog of repair time that accumulates by the time an emergency event occurs, necessitating return to port or tender.

Note that the occurrence rate of emergency events λ can to some degree be controlled by logistics (spares availability) and repair personnel policy. Since unit availability is affected by λ, α, and $E[R]$, the expected repair time for secondary failures, an opportunity exists for an overall optimization that includes tender sojourns. Rather than pursue this further we turn to another model.

Model 3. Let us suppose that m failure-prone systems are cared for by r repairmen, or teams. Suppose, too, that each system fails in Poissonian fashion at rate $\lambda(t)$, where t is measured from time of departure on a mission; likewise repairs are completed at rate $\mu(t)$. The time dependence of $\lambda(t)$ and $\mu(t)$ may be introduced to represent a learning effect: perhaps $\lambda(t)$ decreases with t, and $\mu(t)$ increases with time t to reflect training and leadership improvements. Now, with the aid of independence assumptions it may be shown that $\{N(t), t \geq 0\}$, the number of systems down and awaiting maintenance at t, is a birth and death Markov process. Usually, however, simple explicit representations for $P\{N(t) = j \mid N(0) = i\}$, $E[N(t)]$, and more interesting figures of merit are difficult to come by.

It has been shown by Iglehart [12.6], using results of C. Stone on weak convergence, that if m is large a properly normalized version of $N(t)$ may be approximated by a diffusion, Ornstein-Uhlenbeck (O.U.) process. Further developments were given by Schach and McNeill [12.11], and by McNeill [12.12]. We produce a simple heuristic derivation, and then suggest some useful calculations and further models.

Let $dN(t)$ denote the change in $N(t)$ in time dt. This may be thought of as a deterministic, average, movement plus a random element:

$$dN(t) = \lambda(t)[m-N(t)]dt = N(t)\mu(t)dt + \sqrt{\lambda(t)[m-N(t)]+N(t)\mu(t)}\, dW(t), \qquad (12.24)$$

if $N(t) \leq r$

while

$$dN(t) = \lambda(t)[m-N(t)]dt - r\mu(t)dt$$

$$+ \sqrt{\lambda(t)[m-N(t)]+r\mu(t)}\, dW(t) ,$$

$$\text{if } N(t) > r$$

In (12.24) the right most term is a scale factor that is proportional to the standard deviation of the difference of two Poisson processes operating over $(t,t+dt)$ multiplied by $dW(t)$, the increment of a Wiener process, the latter having mean zero and variance dt. We have thus replaced the actual "difference of Poissons" random component by a Gaussian component. The latter is certainly plausible in the limit as large m makes the Poisson arrival rate large.

Now, if $r = \infty$ (infinitely many servers), the failures are repaired without delay, and $N(t)$ has a Poisson distribution. The latter approaches a Gaussian form as $m \to \infty$. Consequently, we expect that if $r = cm$ we can derive a Gaussian form approximating $N(t)$. To this end, put

$$S(t) = \frac{N(t) - mx(t)}{\sqrt{m}}$$

or

$$N(t) = mx(t) + \sqrt{m}\, S(t)$$

where $x(t)$ is a deterministic part, and $S(t)$ a random noise; the latter is assumed to be bounded with probability one. Then formally

$$dN(t) = mdx(t) + \sqrt{m}\, dS(t) = \lambda(t)[m(1-x(t))$$

$$- \sqrt{m}\, S(t)]dt - \mu(t)[mx(t) + \sqrt{m}\, S(t)]dt \qquad (12.25)$$

$$+ \sqrt{\mu(t)[m(1-x(t)-\sqrt{m}\, S(t)]+\mu(t)[mx(t)=\sqrt{m}\, S(t)]}\, dW(t)$$

$$\text{if} \quad mx(t) + \sqrt{m}\, S(t) \leq r = cm$$

and

$$dN(t) = m\,dx(t) + \sqrt{m}\, dS(t) = \lambda(t)[m(1-x(t)) - \sqrt{m}\, S(t)]dt - \mu(t)cm\, dt + \sqrt{\lambda(t)[m(1-x(t))-\sqrt{m}\, S(t)] + \mu(t)cm}\; dW(t) \,, \tag{12.26}$$

$$\text{if} \quad mx(t) + \sqrt{m}\, S(t) > r = cm$$

Next, divide throughout by $\sqrt{m}$ and let $m \to \infty$. We find that in order for equality to hold in (12.25) and (12.26),

$$\frac{dx}{dt} = \lambda(t)[1-x(t)] - \mu(t)x(t) \,, \quad \text{if} \quad x(t) \leq c$$

$$\frac{dx}{dt} = \lambda(t)[1-x(t)] - \mu(t)c \,, \quad \text{if} \quad x(t) > c \tag{12.27}$$

The solution of these equations describes the mean motion of the backlogged repairs according to our approximation. In particular, if $\lambda(t) = \lambda$, $\mu(t) = \mu$, both constants and $t \to \infty$ we find that

$$x(\infty) = \frac{\lambda}{\lambda+\mu} \,, \quad \text{if} \quad x(\infty) = \frac{\lambda}{\lambda+\mu} < c$$

while

$$x(\infty) = 1 - \frac{\mu}{\lambda}\, c \,, \quad \text{if} \quad x(\infty) > c \quad \text{and} \quad \mu c < \lambda$$

Next, the terms of order $\sqrt{m}$ remain; after taking limits inside the scale factor term, recalling that $S(t)$ is bounded in probability, we find that

$$dS(t) = -[\lambda(t)+\mu(t)]S(t)dt +$$

$$+ \sqrt{\lambda(t)[1-x(t)]+\mu(t)x(t)}\ dW(t)\ , \qquad (12.28a)$$

$$\text{if}\quad x(t) < c$$

and

$$dS(t) = -\lambda(t)S(t)dt + \sqrt{\lambda(t)[1-x(t)]+\mu(t)c}\ dW(t)\ , \qquad (12.28b)$$

$$\text{if}\quad x(t) > c$$

Our conclusion is that the limiting $(m\to\infty)$ noise process $\{S(t), t \geq 1\}$ is a nonstationary O.U. diffusion; however, one such O.U. diffusion, (12.28a), describes the noise for $x(t) < c$, and a different one, (12.28b) $x(t) > c$. In passing we note that if $\lambda(t) = \lambda$, $\mu(t) = \mu$, and the time scale change $\tau = (\lambda+\mu)t$ is made that then (12.28a) is directly solved to give

$$S(\tau) = S(0)e^{-\tau} + \int_0^\tau e^{-(\tau-x)}dW(\tau)$$

indicating that the noise is Gaussian with limiting variance $\lambda\mu/(\lambda+\mu)^2$--entirely in conformity with the results of Iglehart and Lemoine [12.7]; limiting results for (12.28b) will also agree with those of [12.7]. While definite conditions must be imposed on $\lambda(t)$ and $\mu(t)$ to validate our derivations, the results obtained seem to be fully in agreement with those obtained by careful weak convergence arguments. Of equal importance in the applications is the quality of the approximations obtained when the number of systems m and repairmen r is finite. Numerical and simulational comparisons are under way at the Naval Postgraduate School to check (and improve) the quality of the approximation.

One obvious use of the above approximations is that of representing system failure status at t, t being measured from start of a mission. Clearly the expected number of failures awaiting repair is, if $N(0) = 0$,

$$E[N(t)] \tilde{\approx} mx(t)$$

$x(t)$ being the solution of (12.27); the variance may be found from (12.28a,b). These approximations allow an assessment of system status at any fixed time t. Moreover, at least in the case of constant λ and μ we can, using tables of Keilson and Ross [12.8], assess the probability that $N(t) \tilde{<} mx(t) + \sqrt{m}\, S(t) \leq B(t)$ for all t, $0 \leq t \leq T$. Other functionals, for example, total accumulated time spent waiting for repair, are similarly accessible; such is not usually true when birth and death formulations are utilized in the manner of Feller [12.1] (1,Chapter 17).

The modeling procedure described can also be applied to situations in which overall repair rate has general state dependency, representing a sort of priority assignment. It may also be used to represent the consequences of simultaneous, shock-model type failures. Owing to the relative simplicity of the solutions, one can contemplate an assignment of maintenance and repair facilities and personnel so as to optimize various meaningful measures of effectiveness subject to desired constraints.

12.6 Conclusions

Although a variety of probabilistic methods useful in logistics studies have been illustrated, the historical accounting is far from complete. An informal list of omissions should include (a) classical inventory theory (Iglehart, Veinott, Arrow, Karlin, Scarf, Harris and Gross, and so on), (b) sequential Bayesian methods (Chernoff, De Groot, and so on), probabilistic mathematical programming and stochastic control theory (many authors), and manpower modeling (Oliver, Marshall, Grinold). Finally, it seems worth noting that under present circumstances much opportunity exists for problem formulation and solution in areas that include both manpower supply and allocation and diagnostics, maintenance, and repair.

References

[12.1] Feller, W. (1968 and 1971). An Introduction to Probability Theory and its Applications. 1 and 2. Wiley.

[12.2] Gaver, D. P. (1968). Diffusion approximations

and models for certain congestion problems. J. Appl. Probability 5 607-623.

[12.3] Gaver, D. P., J. P. Lehoczky, and M. Perlas (1975). Service systems with transitory demand. in M. A. Geisler (ed.) Logistics. North-Holland/TIMS Studies in the Management Sciences 1 21-34.

[12.4] Gnedenko, B. V. (1967). Some theorems on standbys. in L. M. LeCam and J. Neyman (eds.) Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability III. University of California. 285-291.

[12.5] Howard, R. A. (1971). Dynamic Probabilistic Systems. 1 and 2. Wiley.

[12.6] Iglehart, D. L. (1965). Limiting diffusion approximations for the many server queue and the repairman problem. J. Appl. Probability 2 429-441.

[12.7] Iglehart, D. L., and A. J. Lemoine (1973). Approximations for the repairman problem with two repair facilities, I: no spares. Advances in Appl. Probability 5 595-613.

[12.8] Keilson, J., and H. Ross (1971). Passage time distributions for the Ornstein-Uhlenbeck process. Preliminary report, Statistics Department, University of Rochester (September).

[12.9] Lewis, P. A. W. (ed.) (1972). Stochastic Point Processes: Statistical Analysis, Theory and Applications. Wiley.

[12.10] Luckew, R. S. (1971). Generalized alternating renewal process models. Technical Report No. 51, Department of Statistics, Carnegie-Mellon University (June).

[12.11] McNeill, D. R., and S. Schach (1973). Central limit analogues for Markov population processes (with discussion). J. Royal Statist. Soc. Ser. B 35 1-23.

[12.12] McNeill, D. R. (1973). Diffusion limits for congestion models. J. Appl. Probability 10 368-376.

Chapter 13

RECENT RESEARCH ON CLASSES OF LIFE DISTRIBUTIONS USEFUL IN MAINTENANCE MODELING*

Frank Proschan
The Florida State University

13.1 Introduction

In this chapter we shall survey some classes of life distributions recently introduced for use in the study of maintenance policies. We consider mainly two types of planned maintenance policies.

Definition. Under an age replacement policy, a unit is replaced upon failure or at age T, whichever comes first. (13.1)

Definition. Under a block replacement policy, the unit in operation is replaced upon failure and at times $T, 2T, 3T, \ldots$. (13.2)

An age replacement policy is more difficult and costly to administer since it requires keeping track of the age of the device. However, it does have the advantage that fewer relatively new items are replaced. Block replacement is generally used in the maintenance of a "block" of similar items, such as the set of neon tubes used on a given floor of an office building. Minimal record-keeping is involved, since all tubes are replaced at regular intervals, and in addition, failed tubes are replaced at failure. Intuitively, it seems clear that block replacement will lead to replacement of a greater number of unfailed items. In addition to the above two planned replacement policies, we shall consider the policy defined in

Definition. Under a replace at failure only policy, the unit is replaced only when it fails. (13.3)

*The preparation of this chapter was supported by the Air Force Office of Scientific Research under Grant AFOSR 74-2581.

Note that the sequence of failure intervals corresponds to a renewal process.

In Section 13.2, we compare stochastically the three replacement policies as to number of failures during operation, number of planned replacements, total number of removals, and so on. We find that certain classes of life distributions arise naturally and play a crucial role in the comparison and study of these maintenance policies.

Definition. A life distribution F (or survival function $\bar{F} \stackrel{def}{=} 1-F$) is said to be New Better than Used (NBU) if

$$\bar{F}(x+y) \leq \bar{F}(x)\bar{F}(y) \quad \text{for all} \quad x,y \geq 0 \tag{13.4}$$

This means that for each $x > 0$, the probability $\bar{F}(x)$ that a new item survives a period of length x is greater than the corresponding probability that an unfailed item of age y survives an additional period of length x. Another way of stating this is that a used item has stochastically smaller remaining life length than does a new item. Mathematically, (13.4) states that $-\log \bar{F}(t)$ is superadditive.

Definition. A life distribution F (or survival function $\bar{F}$) is said to be New Better than Used in Expectation (NBUE) if the mean μ of F is finite and

$$\int_0^\infty [\bar{F}(t+x)/\bar{F}(t)]dx \leq \mu \tag{13.5}$$

for all $t \geq 0$ such that $\bar{F}(t) > 0$.

Note that $\int_0^\infty [F(t+x)/\bar{F}(t)]dx$ represents the conditional mean remaining life of a unit of age t; hence the inequality states that a used unit of age t has smaller mean remaining life than a new unit if F is NBUE.

By reversing the direction of inequality for $\bar{F}$ in (13.4) and for the integral in (13.5), respectively, we obtain dual classes, New Worse than Used (NWU) and New Worse than Used in Expectation (NWUE). The results stated below for NBU and NBUE have obvious duals for NWU and NWUE; these dual results will not be explicitly stated.

In our discussion of the application of the NBU, NBUE, NWU, and NWUE classes in the study of maintenance policies, we will find it helpful to recall the definitions of other classes of life distributions, some of which have already played a significant role in reliability and life testing.

Definitions. A distribution function F or survival function $\bar{F}$ is said to be or to have (13.6)

(i) Increasing Failure Rate (IFR) if $\bar{F}(x+t)/\bar{F}(t)$ is decreasing in t whenever $x > 0$

(ii) Decreasing Mean Residual Life (DMRL) if $\int_0^\infty [\bar{F}(x+t)/\bar{F}(t)]dx$ is decreasing in t

(iii) Increasing Failure Rate Average (IFRA) if $[\bar{F}(t)]^{1/t}$ is decreasing in $t > 0$

When F has a density, (i) is equivalent to the condition that for some version f of the density, the hazard rate $r(t) \stackrel{\text{def}}{=} f(t)/\bar{F}(t)$ is increasing in t. Also F is IFR if and only if $\log \bar{F}$ is concave. To say that the life distribution F of an item is IFR is to say that the residual life length of an unfailed item of age t is stochastically decreasing in t. To say that the life distribution F of an item is DMRL is equivalent to saying that the residual life of an unfailed item of age t has a mean that is decreasing in t. When a failure rate exists, (iii) is equivalent to the condition that the failure rate average $t^{-1}\int_0^t r(u)du$ is increasing in t.

Equivalently, (iii) also says that $-t^{-1}\log\bar{F}(t)$ is increasing in t, that is, $-\log\bar{F}(t)$ is a star-shaped function.

Dual classes are obtained by replacing "decreasing" by "increasing" and "increasing" by "decreasing" in (13.6) (i), (ii), and (iii); these classes are called respectively, Decreasing Failure Rate (DFR), Increasing Mean Residual Life (IMRL), and Decreasing Failure Rate Average (DFRA). The inclusion relations among the classes defined above may be graphically displayed as shown in Figure 13.1. See, for example, Bryson and Siddiqui (1969). An additional class of distributions has recently been introduced by Haines and Singpurwalla (1974); it is defined and discussed in Section 13.6.

13.2 Replacement Policy Comparisons

A major purpose of planned replacement policies is to minimize the probability of failure during operation. Thus, it is of importance to compare stochastically the numbers of failures observed under the three types of maintenance policies described in Definitions (13.1,

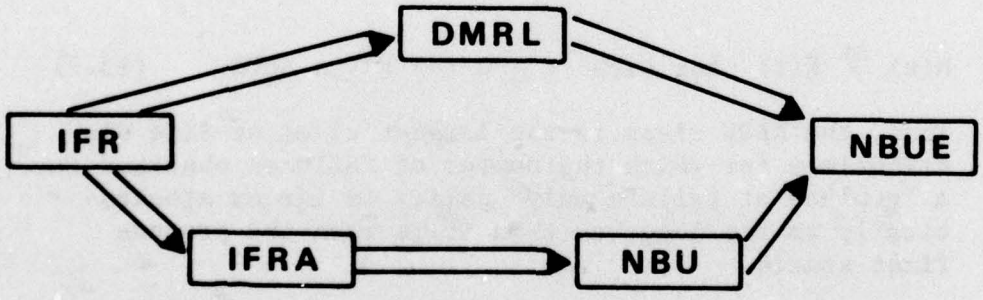


Figure 13.1 - Inclusion relations among classes of life distributions.

.2, and .3). Recall that random variable X is stochastically larger than random variable Y, written $X \overset{st}{>} Y$, if $P[X > x] > P[Y > x]$ for each real x. Let us also adopt the following notation.

$N(t)$ = the number of failures during $[0,t]$ under a "replace at failure only" policy; that is, it equals the number of renewals in an ordinary renewal process

$\hat{N}(t)$ = the number of renewals during $[0,t]$ in a stationary renewal process, that is, in a renewal process that starts at time $-\infty$

$N_A(t,T)$ = the number of failures during $[0,t]$ using an age replacement policy with planned replacement age T

$N_B(t,T)$ = the number of failures during $[0,t]$ using a block replacement policy with planned replacement interval T

Marshall and Proschan (1972) prove the following results.

$$\hat{N}(t) \overset{st}{>} N(t) \text{ for each } t \geq 0 \iff F \text{ is NBUE.} \tag{13.7}$$

Thus, the NBUE class is the largest class of life distributions for which the number of failures observed in a "replace at failure only" policy is larger stochastically in the long run than it is when the process first starts.

$$N(t) \overset{st}{>} N_A(t,T) \text{ for each } t \geq 0, T \geq 0 \iff F \text{ is NBU.} \tag{13.8}$$

Consequently, the NBU class is the largest class of life distributions for which age replacement stochastically decreases the number of failures. It follows that the NBU class is the natural class of life distributions in the study of age replacement policies. A

similar result holds for block replacement policies.

$$N(t) \overset{st}{>} N_B(t,T) \quad \text{for each} \quad t \geq 0 \ , \ T \geq 0 \iff F \text{ is NBU.} \tag{13.9}$$

Next we can compare block replacement policies calling for different intervals between planned replacements.

$$N_B(t,kT) \overset{st}{>} N_B(t,T) \quad \text{for each} \quad t \geq 0 \ , \ T \geq 0 \ , \ k = 1,2,\ldots \iff F \text{ is NBU.} \tag{13.10}$$

Therefore, the NBU class is the largest class of life distributions for which a planned replacement interval length of an integer multiple of T, as compared with one of length T, results in more failures stochastically. A similar result holds for age replacement.

$$N_A(t,kT) \overset{st}{>} N_A(t,T) \quad \text{for each} \quad t \geq 0 \ , \ T \geq 0 \ , \ k = 1,2,\ldots \iff F \text{ is NBU.} \tag{13.11}$$

A finer comparison is possible if we confine ourselves to the more highly structured class of IFR distributions.

$$N_A(t,T_1) \overset{st}{<} N_A(t,T_2) \quad \text{for each} \quad T_1 < T_2 \ , \ t \geq 0 \iff F \text{ is IFR.} \tag{13.12}$$

Note that this stochastic comparison no longer requires an integer multiple of planned replacement interval T.

Up till now, our comparisons have been confined within each class of replacement policies. Next we compare age versus block replacement policies. Let $R_A(t,T)(R_B(t,T))$ denote the number of removals during $[0,t]$, including both failed and unfailed items, using an age (block) replacement policy with planned replacement interval T. Barlow and Proschan (1964) prove the following.

For each $t > 0$, $T > 0$,

For every life distribution,

$$R_A(t,T) \overset{st}{<} R_B(t,T) \tag{13.13}$$

For an underlying IFR life distribution,

$$N_A(t,T) \overset{st}{>} N_B(t,T) \tag{13.14}$$

13.3 Models for the NBU and NBUE Classes

A number of physically motivated models have been proposed that yield the NBU and NBUE classes of life distributions.

Coherent systems of repairable components. Ross (1974) considers a coherent system of components. (See Barlow and Proschan (1975) *Chapters 1 and 2, for a discussion* of coherent systems.) Component i has exponential life length with failure rate λ_i ; upon failure, repair is initiated requiring exponential repair period with repair rate ν_i . All life lengths and repair periods are mutually independent. Ross proves that the time to the first system failure has an NBU distribution. He further notes, by contrast, the interval of time between later successive system failures need not be NBU.

Shock models. Esary, Marshall, and Proschan (1973) consider a device subject to shocks occurring in time according to a Poisson process with shock rate λ . The probability that the device survives k shocks is $\bar{P}_k$, where $1 = \bar{P}_0 \geq \bar{P}_1 \geq \bar{P}_2 \geq \cdots$. Then the survival probability $\bar{H}(t)$ over time is given by

$$\bar{H}(t) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \bar{P}_k \quad \text{for} \quad t \geq 0 \tag{13.15}$$

Esary, Marshall, and Proschan obtain results in which various classes of life distributions are "preserved"

under the transformation (13.15). Specifically, they prove the following.

Theorem 1 (a) If $\bar{P}_k$ is discrete NBU (that is, $\bar{P}_{k+\ell} \leq \bar{P}_k \bar{P}_\ell$ for each $k = 0,1,2,\ldots$; $\ell = 0,1,2,\ldots$), then $\bar{H}(t)$ is NBU.

(b) If $\bar{P}_k$ is discrete NBUE (that is, $\sum_{j=0}^{\infty} \bar{P}_j \geq \sum_{j=0}^{\infty} (\bar{P}_{k+j}/\bar{P}_k)$ for $k = 0,1,2,\ldots$) , then $\bar{H}(t)$ is NBUE.

A-Hameed and Proschan (1973, 1975) obtain similar preservation results in the following more general models.

(1) Shocks occur according to a nonstationary Poisson process

(2) Shocks occur according to a birth process

Preservation of classes of life distributions. An important question in formulating classes of life distributions is the following. For which standard reliability operations is the class of life distributions closed? For example, is the convolution of NBU distributions itself NBU? Note that the convolution of distributions corresponds to the addition of independent life lengths; such an operation arises routinely in the determination of spares kits. We summarize the situation in Table 13.1. The operation "formation of coherent systems" refers to the situation in which a coherent system is formed of independent components, not subject to repair. An arbitrary mixture F of distributions $F_1,\ldots,F_n$ is given by

$$F = p_1F_1 + \cdots + p_nF_n$$

where each $p_i \geq 0$ and $\sum_1^n p_i = 1$. Finally,

Table 13.1 Preservation of Life Distribution Classes Under Reliability Operations

Class	Formation of Coherent Systems	Convolutions
NBU	Preserved	Preserved
NBUE	Not Preserved	Preserved
NWU	Not Preserved	Not Preserved
NWUE	Not Preserved	Not Preserved

Class	Arbitrary Mixtures	Mixtures of Distributions That Do Not Cross
NBU	Not Preserved	Not Preserved
NBUE	Not Preserved	Not Preserved
NWU	Not Preserved	Preserved
NWUE	?	Preserved

distributions F_1 and F_2 are said not to cross if there is no pair t_1, t_2 such that $F_1(t_1) - F_2(t_1) > 0$ and $F_1(t_2) - F_2(t_2) < 0$. The preservation results summarized in Table 13.1 are proved in Marshall and Proschan (1972) and Esary, Marshall, and Proschan (1970).

13.4 Bounds for the NBU and NBUE Classes

In this section we survey results on bounds for individual components and bounds on system mean life.

Bounds for individual components. Marshall and Proschan (1972) develop the following simple bound for the NBU distribution.

Theorem 2. Let F be NBUE with mean μ. Then

$$F(t) \leq t/\mu \quad \text{for} \quad t \leq \mu$$

and the bound is sharp.

It is interesting to note that the above bound cannot be improved even if F is restricted to the smaller NBU class. Haines and Singpurwalla (1974) do obtain a stronger bound for the NBUE class by assuming additional information is known, as stated in the following.

Theorem 3. Let F be NBUE with mean μ and $\bar{F}(t_0) = \alpha$ for some $0 \le t_0 \le \mu$. Then

$$\bar{F}(t) \ge \text{Max}\left(\alpha, \frac{\mu - t}{\mu}\right) \quad \text{for} \quad 0 \le t \le t_0$$

$$\ge \frac{1}{\mu}\left[\mu - t_0 - \alpha(t - t_0)\right] \quad \text{for} \quad 0 \le t_0 \le t \le \beta$$

where $\beta = \dfrac{\mu + t_0\alpha - t_0}{\alpha}$.

For the NBU class, Marshall and Proschan (1972) present the following bounds.

Theorem 4. Let F be NBU and $\bar{F}(t_0) = \alpha$. Then

$$\bar{F}(t) \ge \alpha^{1/k} \quad \text{for} \quad \frac{t_0}{k+1} < t \le \frac{t_0}{k} \quad \text{and} \quad k = 1,2,\ldots$$

while

$$\bar{F}(t) \le \alpha^{k} \quad \text{for} \quad kt_0 \le t < (k+1)t \quad \text{and} \quad k = 0,1,2,\ldots$$

and these bounds are sharp.

The upper bound is itself an NBU survival function; the lower bound is not.

Bounds on system mean life. The bounds given above have been for individual NBU and NBUE components. Marshall and Proschan (1970, 1972) give bounds for the mean life of series and parallel systems of NBUE components.

Theorem 5. Let $\mu_s(\mu_p)$ be the mean life of a series

(parallel) system of n independent NBUE components with mean lives respectively of $\mu_1,\ldots,\mu_n$. Then

$$\mu_s \geq \left(\sum_1^n \mu_i^{-1}\right)^{-1}$$

$$\mu_p \leq \int_0^\infty \left[1 - \prod_1^n \left(1 - e^{-x/\mu_i}\right)\right] dx$$

and the bounds are sharp.

Note that the bounds represent, respectively, the mean life of a series and parallel system of exponential components.

13.5 Statistical Inference

Given a sample $X_1,\ldots,X_n$ from life distribution F, Hollander and Proschan (1972) propose a test of the hypothesis

H_0: The distribution F is exponential with unspecified scale parameter versus the alternative hypothesis

H_1: The distribution F is NBU (and not exponential)

The test statistic proposed is motivated by consideration of the parameter

$$\gamma(F) \stackrel{\text{def}}{=} \iint [\bar{F}(x)\bar{F}(y) - \bar{F}(x+y)]dF(x)dF(y) \qquad (13.16)$$

Note that the integrand is identically 0 when F is exponential, and nonnegative when F is NBU. The test statistic is developed by first replacing in (13.16) the unknown distribution F by the empirical distribution F_n. Next the U-statistic which is asymptotically equivalent is used, since U-statistics have many desirable properties and a fully developed theory. The test

statistic essentially counts the number of triples of ordered observations $X_{(i)} < X_{(j)} < X_{(k)}$ such that $X_{(i)} + X_{(j)} > X_{(k)}$ and rejects for large values. The statistic is unbiased, asymptotically normal, and is consistent. The asymptotic relative efficiency of the test statistic is determined relative to statistics designed against IFR alternatives (since no other test statistics have yet been proposed against NBU alternatives). The NBU test statistic proposed shows reasonably good asymptotic relative efficiency, especially if one takes into account the fact that the class of IFR alternatives is more restricted than is the class of NBU alternatives. To permit application of the test, small sample null tail probabilities are derived, and additional critical values are obtained by Monte Carlo sampling. Tables are provided for sample sizes up to 50. For larger sample sizes, the asymptotic normality may be used, along with the calculated asymptotic mean and variance.

Other than the test for NBU, statistical inference for the NBU, NBUE, NWU, and NWUE has not been developed as yet. A useful contribution would be to develop an estimator with desirable properties for, say, the NBU class of distributions.

13.6 Related Classes of Life Distributions

Several other classes of life distributions have been proposed for use in the study of maintenance policies. In certain respects, they seem less appropriate for maintenance modeling than do the NBU and NBUE classes and their duals, discussed above. However, we summarize some of the recent work on these classes. We consider in particular the DMRL class, defined in (13.6), and the following new class introduced and studied by Haines and Singpurwalla (1974).

Definition. A distribution F has Decreasing Percentile Residual Life (DPL) if for some α, $0 < \alpha < 1$, the 100α-th percentile of the residual life of an item of age t decreases in $t \geq 0$. (13.17)

A DPL distribution is somewhat similar to a DMRL

distribution in that in both cases a parameter measuring residual life is a decreasing function of age. Haines and Singpurwalla (1974) develop a number of properties of the DPL class and relate it to classes of life distributions developed earlier. A typical result is the following.

Theorem 6. An IFR distribution is DPL for each value of α in $(0,1)$.

One seeming disadvantage of the DPL class is that except for the inclusion stated in Theorem 6, the DPL class neither contains nor is contained in any of the other classes of life distributions. Another somewhat discouraging aspect of the DPL class is that none of the preservation properties shown in Table 13.1 can be claimed for either the DPL or its dual class.

Haines and Singpurwalla obtain bounds on the survival function of both the DPL and the DMRL distribution functions having a known mean and a percentile. They present a series of graphs portraying various bounds for the various classes of distributions; the graphs show how the bound improves as additional information is used.

Finally, we mention an interesting display of the empirical mean residual life for a group of cancer patients, shown in Bryson and Siddiqui (1969). The graph shows quite dramatically that the mean residual life length is decreasing in time measured from diagnosis of cancer; that is, the population is DMRL. It would be desirable to have an estimator for the DMRL class having optimal properties.

References

[13.1] Barlow, R. E., and F. Proschan (1964). Comparison of replacement policies, with renewal theory implications. Ann. Math. Statist. 35 577-589.

[13.2] Barlow, R. E., and F. Proschan (1975). Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart, and Winston.

[13.3] Bryson, M., and M. Siddiqui (1969). Some criteria for aging. J. Amer. Statist. Assoc. 64

1472-1483.

[13.4] Esary, J. D., A. W. Marshall, and F. Proschan (1970). Some reliability applications of the hazard transform. SIAM J. Appl. Math. 18 849-860.

[13.5] Esary, J. D., A. W. Marshall, and F. Proschan (1973). Shock models and wear processes. Ann. of Probability 1 627-649.

[13.6] Haines, A. L., and N. D. Singpurwalla (1974). Some contributions to the stochastic characterization of wear. in F. Proschan and R. J. Serfling (eds.) Reliability and Biometry. Society for Industrial and Applied Mathematics. 47-80.

[13.7] A-Hameed, M. S., and F. Proschan (1973). Nonstationary shock models. Stochastic Processes Appl. 1 383-404.

[13.8] A-Hameed, M. S., and F. Proschan (1975). Shock models with underlying birth process. J. Appl. Probability 12 18-28.

[13.9] Hollander, M., and F. Proschan (1972). Testing whether new is better than used. Ann. Math. Statist. 43 1136-1146.

[13.10] Marshall, A. W., and F. Proschan (1970). Mean Life of series and parallel systems. J. Appl. Probability 7 165-174.

[13.11] Marshall, A. W., and F. Proschan (1972). Classes of distributions applicable in replacement, with renewal theory implications. in L. M. LeCam, J. Neyman, and E. L. Scott (eds.) Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability I. University of California. 395-415.

[13.12] Ross, S. (1974). On time to first failure in multicomponent exponential reliability systems. Operations Research Center Report ORC 74-8, University of California, Berkeley. To appear in Stochastic Processes Appl.

Part V

MATHEMATICAL PROGRAMMING

Chapter 14

EXTREMAL METHODS IN LOGISTICS RESEARCH: A DEVELOPMENTAL SURVEY*

A. Charnes
The University of Texas
W. W. Cooper
Carnegie-Mellon University
Edward S. Bres III
The University of Texas

14.1 Introduction

The inventory in this chapter represents a compilation from the articles published in the Naval Research Logistics Quarterly which we hereafter refer to as the Quarterly. Our list is by no means exhaustive but it does contain over 275 titles covering 20 years of publications in one of the most prominent sources of results from logistics research. We hope that our list will provide a convenient source for examining some of the uses of extremal methods in logistics and that it will portray research aspects and developments in a relatively prominent form.

At the outset we should indicate that "extremization" or an "extremal method" refers to the optimizing operators such as "max" or "min" or "min-max" or "max-min," and so on. It also extends to the operations such as "inf" or "sup" where a direction of optimization is indicated. Finally, an "extremal equation" refers to any equation whose elements are required to satisfy prescribed extremization conditions or operations. This is by analogy, or extension, of the way one characterizes differential equations or integral equations in terms of differentiation or integration operations.

A check of the articles surveyed below will make it apparent that there has been a vigorous and "varied" growth in the development of extremal methods. Ideally,

*The preparation of this chapter was supported by the Office of Naval Research under Contracts N00014-67-A-0126-0008 and N00014-67-A-0126-0009 with the University of Texas.

we would have liked to match these methods against significant classes of problems to portray their potential (or actual) uses and shortcomings. At a minimum, we would have liked to arrange a two-way classification or matrix array in which we could have at least indicated where uses of linear programming, game theory, or other such developments embodying extremal methods or approaches had been used on classes of logistics problems such as manpower planning or fleet provisioning with further breakdowns into recruitment and procurement, storage and transfer, and so on. The format of the articles and the arrangements in the Quarterly, however, prevented us from doing this. But we were able to make a rough classification of these articles by methods and problems addressed and, in the next section, we discuss specific strategies for applying extremal methods to logistics problems.

14.2 Strategies for Applications

In order to enhance the usefulness of our survey, and to provide perspective, we will discuss strategies of the kinds we have ourselves used in bringing extremal methods to bear on logistics problems. We will proceed in a somewhat informally structured way, using four questions.

What are Some of the Rationales for the Use of Extremal Methods? We have become so accustomed to discussions about the relevance or realism of particular objectives of optimization in game theory, economic theory, and so on, that we can fail to remember that optimizations can also be used as artifacts to simplify an otherwise complex analysis or diminish ambiguity, and so on. We have discussed these uses elsewhere in considerable detail[1] and hence we simply indicate below some of the related further considerations.

(1) Representation in extremization form can provide ease in formulating problems and assessing potential

[1] See A. Charnes and W. W. Cooper (1963). A strategy for making models in linear programming. in R. Machol, et al. (eds.) Systems Engineering Handbook. McGraw-Hill.

further consequences. An example would be game theory as used in representing and studying traffic flow problems.[2]

(2) Extremal methods can provide ruggedness in the face of data deficiencies and can supply computational power as required for operational implementation. The example just supplied becomes a case in point when the already achieved game theoretic formulation for traffic flow is transformed into a corresponding linear program in order to achieve computational power. The associated optimization may then also be arranged so that the results are not responsive to minor data variations, or even to marked data variations that lie away from the frontier on which an optimum occurs. This illustrates what we refer to as the property of ruggedness associated with such optimizations.

(3) Extremal principles can facilitate and provide access to interpretations that might otherwise not be possible. An example would be optimization of a traffic flow problem to provide access to equilibrium characterizations that would not have been apparent from other approaches such as unguided simulation. Such optimizations as in, for example, Pareto types of optimality,[3] can also supply clues to underlying behavior, and hence also provide guides to system redesigns when, for instance, alternate routes are to be opened to drivers in search of origin-to-destination times.

How Can One or More Extremal Methods be Evaluated for a Particular Application? This question is, of course, not independent of the preceding but let us use it to concentrate on "flexibility and ease of extension."

---

[2] See A. Charnes and W. W. Cooper (1961). Management Models and Industrial Applications of Linear Programming. II Wiley. Chapter XX.

[3] More precisely, the reference is to Wardrop's second principle as discussed in Charnes and Cooper (1961), Section XX.5.

Specifically, extremal methods can be evaluated on the following bases.

(1) A method can involve useful alternative representations. One example is linear programming and the way it provides access to a dual problem whose solution at an optimum may be used to study the original problem. A second example is that recourse to an appropriate extremal model may be used to guide or control a simulation that, lacking such a guide, may be inefficient and ambiguous as well. See, for example, the preceding discussion of traffic flow models.

(2) An extremal method can involve alternative methods or mechanisms. An example would be a decomposition for computation and also for control of decentralized operations by placing bounds or providing operational guides. This can be regarded as a computational expedient, as when the optimization of a large problem is replaced by optimizing several smaller ones. It can also supply insight into new modes of operation or ways of improving existing modes as when, say, feedback mechanisms are utilized to alter data supplied for guidance in the decisions of decentralized divisional managers.

How Can Extremal Methods be Justified? Again, there is overlap with our other questions but here we particularly address problems of convincing managers.

(1) An extremal method may be distinguished by the ease of explanation of its managerial uses. For example, this should be the case for a tradeoff analysis where assurance can be given that there are requisite (that is, extremal) tradeoff terms.

(2) A method may constitute a guide for further scientific research as when, say, the deeper analysis provided by an extremizing characterization ties together and reveals relations between many seemingly

disparate problems or approaches to problems.[4]

How Can Extremal Methods Provide Entrance to New Methods or New Problems? Examples of the former were touched upon under preceding questions. Here we wish to discuss how new methods, and especially methods involving extremal approaches, lead to new problems; and conversely, there is the possibility that as one approaches new problems, or refinements of old ones, it may become evident where and in what forms new methods or combinations of preexisting methods are wanted.

An example from some of our own recent experiences with developing manpower planning models in conjunction with R. J. Niehaus and others of the U. S. Navy[5] may be used for illustration. In response to this problem, the state of the modeling art was altered by including, for the first time, Markovian elements in a goal-programming format. Next, the customary divisions between "manpower planning" and "career analysis" were seen to be capable of simultaneous rather than separate treatment. This led to a reconsideration of the ideas of organization design in a new dynamic format. Thus, new possibilities for future research (for example, in organization theory) were identified, as were areas for possible new applications. The work undertaken with D. Cass and others at the Bureau of Naval Personnel provides an example.[6] This work, which was directed to extending the consideration of Navy officer rotations

---

[4]See A. Charnes, W. W. Cooper, and D. B. Learner (1975). Constrained information theoretic characterizations in consumer purchase behavior, Market Research Corporation of America Report 75-1. Submitted to Journal of Market Research.

[5]See A. Charnes, W. W. Cooper, and R. J. Niehaus (1972). Studies in Manpower Planning. U.S. Navy Office of Civilian Manpower Management. Washington.

[6]See D. Cass, A. Charnes, W. W. Cooper, and R. J. Niehaus (1973). A program for Navy officer distribution models. Submitted to Management Sci.

beyond a single period, also opened the possibility of considering the career profiles associated with each such rotation. This, in turn, identified a need for a model and associated methods to match these possibilities within a multiple-objective optimization--and, naturally, further such possibilities for development will continue to evolve as this work proceeds. This work, along with the work at the U. S. Navy Office of Civilian Manpower Management, also resulted in problem enlargements that made the need for improved algorithms apparent. One set of such developments brought the officer rotation models[7] within range for some of the ultra high-speed algorithms that have been developed for transportation and assignment type models.[8] Another set has yielded improvements in the solution routines for goal programming by virtue of special structures which can be associated with such multiple objective optimizations.[9]

---

[7]See D. Cass, A. Charnes, W. W. Cooper, and R. J. Niehaus (1974). A multi-page goal programming model and algorithm for Navy officer rotations. Submitted to Management Sci.

[8]See, for example, R.S. Barr, F. Glover, and D. Klingman (1972). An improved version of the out-of-kilter method and a comparative study of computer codes. Center for Cybernetic Studies Report C. S. 102, University of Texas, Austin. To appear in Math. Programming. Also see V. Srinivasan and G. L. Thompson (1973). Benefit-cost analyses of coding techniques for the primal transportation algorithm. J. Assoc. Comput. Mach. 20 194-213. (The relative efficacy of these codes has brought the computer usages within the range of interactive managerial-computer uses that would otherwise not have been possible.)

[9]See A. Charnes, W. W. Cooper, D. Klingman, and R. J. Niehaus (1975). Explicit solutions in convex goal programming. To appear in Management Sci. The initial ideas underlying goal programming were, curiously enough, also first set forth in a personnel manpower-planning

Many more such developments can be expected as such applications lend point and relevance to the research, and conversely. We have offered the preceding discussion, and the listings and commentaries that follow, not only for some insight into past history but also as a general guide, and perhaps a stimulus, for future interactions between research in extremal methods and applications in logistics.

14.3 The Survey

What follows is a compilation of articles from the Quarterly that have dealt with extremal principles applied to problems of logistics where the latter is considered in its most general sense. This collection is by no means complete, and classifications are of necessity somewhat arbitrary, but it is hoped that these articles will give some examples of the development of extremal principles and their applications in logistics. We preface the classifications with a few introductory comments and include a few brief descriptions.

Linear Programming Methods. The development of efficient linear programming (LP) methods, paralleled by the rapid growth in computational capacity, has been central to the widespread application of mathematical management tools in the post-World War II era. With these developments, large linear programming models are increasingly amenable to solution. Models of a useful size have become practical as operational tools. Beyond the benefits in linear programming applications, these methods have enabled solution of more general problems. Linear programs are repeatedly solved with nonlinear, integer, and combinatorial algorithms, as well as with game theoretic solution procedures. Existence of efficient LP codes is essential to the success of such approaches. The availability of efficient LP procedures gives a direction of attack for solution of more complex problems. A problem that can be shown equivalent to, or approximated by a linear program, even a large such

---

organization-design context. See A. Charnes, W. W. Cooper and R. O. Ferguson (1954). Optimal estimation of executive compensation by linear programming. Management Sci. 1 138-151.

program, is accessible to solution.

Of the articles listed below, Lemke's [14.11] is of particular importance. His approach from the dual side was a fundamental contribution to linear programming methods. Elements of this approach have recurred frequently in dual and primal-dual algorithms.

[14.1] Alway, G. G. (1962). A triangularization method for computations in linear programming. Naval Res. Logist. Quart. 9 163-180.

[14.2] Beale, E. M. L. (1955). Cycling in the dual simplex algorithm. Same J. 2 269-275.

[14.3] Gass, S., and T. Saaty (1955). The computational algorithm for the parametric objective function. Same J. 2 39-45.

[14.4] Graves, G. W. (1965). A complete constructive algorithm for the general mixed linear programming problem. Same J. 12 1-34.

[14.5] Greenberg, H. (1969). A note on a modified primal-dual algorithm to speed convergence in solving linear programs. Same J. 16 271-273.

[14.6] Hadley, G. F., and M. A. Simonnard (1959). A simplified two-phase technique for the simplex method. Same J. 6 221-226.

[14.7] Harris, M. Y. (1970). A mutual primal-dual linear programming algorithm. Same J. 17 199-206.

[14.8] Hartman, J. K., and L. S. Lasdon (1970). A generalized upper bounding method for doubly coupled linear programs. Same J. 17 411-429.

[14.9] Jacobs, W. (1957). Loss of accuracy in simplex computations. Same J. 4 89-94.

[14.10] Kelley, J. E., Jr. (1957). A threshold method for linear programming. Same J. 4 35-45.

[14.11] Lemke, C. E. (1954). The dual method of

solving the linear programming problem. Same J. 1 36-47.

[14.12] Marshall, K. T., and J. W. Suurballe (1969). A note on cycling in the simplex method. Same J. 16 121-137.

[14.13] Nemhauser, G. L. (1964). Decomposition of linear programs by dynamic programming. Same J. 11 191-195.

[14.14] Thompson, P. M. (1957). Editing large linear programming matrices. Same J. 4 97-100.

[14.15] Wagner, H. M. (1958). The dual simplex algorithm for bounded variables. Same J. 5 257-261.

Theory of Games. Military research and applications have been a major force in the development of the theory of games, subsequent to J. von Neumann and O. Morgenstern's foundation of game theory in 1944. A number of articles dealing with game theory have appeared in the Quarterly. The diversity of the articles listed below is a small indication of the wide scope of possible application of the theory of games. Articles in the Quarterly have dealt with formulation and characterization of specific problems, game theoretic approaches to classes of problems, development and extension of the theory of games, and solution procedures. Examples of current relevance are Blackwell's [14.20] and L. S. Shapley's [14.47]. The theory of vector payoff games is still being developed and the concepts of these earlier articles retain relevance.

[14.16] Agnew, R. A., and R. B. Hempley (1971). Finite statistical games and linear programming. Naval Res. Logist. Quart. 18 99-102.

[14.17] Beale, E. M. L., and G. P. M. Heselden (1962). An approximation method of solving Blotto games. Same J. 9 65-79.

[14.18] Blackett, D. W. (1954). Some Blotto games. Same J. 1 55-60.

[14.19] Blackett, D. W. (1958). Pure strategy solutions of Blotto games. Same J. 5 107-109.

[14.20] Blackwell, D. (1954). On multi-component attrition games. Same J. 1 210-216.

[14.21] Braithwaite, R. B. (1959). A terminating iterative algorithm for solving certain games and related sets of linear equations. Same J. 6 63-74.

[14.22] Charnes, A., and R. G. Schroeder (1967). On some stochastic tactical antisubmarine games. Same J. 14 291-311.

[14.23] Chattopadhyay, R. (1969). Differential game theoretic analysis of a problem of warfare. Same J. 16 435-441.

[14.24] Cohen, N. D. (1966). An attack-defense game with matrix strategies. Same J. 13 391-402.

[14.25] Danskin, J. M. (1954). Fictitious play for continuous games. Same J. 1 313-320.

[14.26] Danskin, J. M. (1964). A game over spaces of probability distributions. Same J. 11 157-189.

[14.27] Davis, M., and M. Maschler (1965). The kernel of a cooperative game. Same J. 12 223-259.

[14.28] Eisenman, R. L. (1966). Alliance games of n-persons. Same J. 13 403-411.

[14.29] Griesmer, J. H., R. E. Levitan, and M. Shubik (1967). Toward a study of bidding processes, part IV; games with unknown costs. Same J. 14 415-433.

[14.30] Griesmer, J. H., and M. Shubik (1963a). Toward a study of bidding processes: some constant-sum games. Same J. 10 11-21.

[14.31] Griesmer, J. H., and M. Shubik (1963b). Toward a study of bidding processes, part II: games with capacity limitations. Same J. 10 151-173.

[14.32] Griesmer, J. H., and M. Shubik (1963c). Toward a study of bidding processes, part III: some special models. Same J. 10 199-217.

[14.33] Hale, J. K., and H. H. Wicke (1957). An application of game theory to special weapons evaluation. Same J. 4 347-356.

[14.34] Hershkowitz, M. (1964). A computational note on von Neumann's algorithm for determining optimal strategy. Same J. 11 75-78.

[14.35] Isbell, J. R., and W. H. Marlow (1956). Attrition games. Same J. 3 71-94.

[14.36] Marchi, E. (1967). Simple stability of general n-person games. Same J. 14 163-171.

[14.37] Maschler, M. (1966). A price leadership method for solving the inspector's non-constant-sum game. Same J. 13 11-33.

[14.38] Maschler, M. (1967). The inspector's non-constant-sum game: its dependence on a system of detectors. Same J. 14 275-290.

[14.39] Moglewer, S., and C. Payne (1970). A game theory approach to logistics allocation. Same J. 17 87-97.

[14.40] Mond, B. (1964). On the direct sum and tensor product of matrix games. Same J. 11 205-215.

[14.41] Morrill, J. E. (1966). One-person games of economic survival. Same J. 13 49-69.

[14.42] Owen, G. (1971). Political games. Same J. 18 345-355.

[14.43] Peleg, B. (1965a). Utility functions of money

for clear games. Same J. 12 57-63.

[14.44] Peleg, B. (1965b). An inductive method for constructing minimal balanced collections of finite sets. Same J. 12 155-162.

[14.45] Pruitt, W. E. (1961). A class of dynamic games. Same J. 8 55-78.

[14.46] Sakaguchi, M. (1962). Pure strategy solutions to Blotto games in closed auction bidding. Same J. 9 253-263.

[14.47] Shapley, L. S. (1959). Equilibrium points in games with vector payoffs. Same J. 6 57-61.

[14.48] Shapley, L. S. (1967). On balanced sets and cores. Same J. 14 453-460.

[14.49] Shubik, M., and G. L. Thompson (1959). Games of economic survival. Same J. 6 111-123.

[14.50] Sweat, C. W. (1968). Adaptive competitive decision in repeated play of a matrix game with uncertain entries. Same J. 15 425-448.

[14.51] Thompson, S. P., and A. J. Ziffer (1959). The watchdog and the burglar. Same J. 6 165-172.

[14.52] Thrall, R. M., and W. F. Lucas (1963). n-person games in partition function form. Same J. 10 281-298.

[14.53] Verhulst, M. (1956). The concept of a mission. Same J. 3 45-57.

[14.54] von Neumann, J. (1954). A numerical method to determine optimal strategy. Same J. 1 109-115.

[14.55] Yasuda, Y. (1970). A note on the core of a cooperative game without side payment. Same J. 17 143-149.

[14.56] Zachrisson, L. E. (1957). A tank duel with

game-theoretic implications. Same J. 4 131-138.

Transportation and Assignment Problems. Transportation problems were among the first linear programs studied: for example, Kantorovich in 1939, Hitchcock in 1941, and Koopmans in 1947. The classical Hitchcock-Koopmans transportation (or distribution) problem, and the assignment problem, are characterized by a special structure but there are many other possible transportation problems. Specialized algorithms for the distribution and assignment problems were developed at an early date, using special structure to advantage for greater efficiency and the ability to handle larger problems than general codes. Examples are Kuhn's [14.72], Charnes and W. W. Cooper's "stepping stone" method of 1954, and G. B. Dantzig's "row-column sum" method of 1951. Articles in the Quarterly have dealt with algorithms for the distribution and assignment problems, extensions of these problems, and formulation of other problems of transportation. An article of importance here is Ford and Fulkerson's [14.61]. Here they extended their network flow algorithm to the important case of the distribution problem where routes have limited capacities. Ford-Fulkerson methods, which rigorized Boldyreff's heuristic "method of flooding," have been a standard approach in distribution and network problems.

[14.57] Beale, E. M. L. (1959). An algorithm for solving the transportation problem when the shipping cost over each route is convex. Naval Res. Logist. Quart. 6 43-56.

[14.58] Bellmore, M., W. D. Eklof, and G. L. Nemhauser (1969). A decomposable transshipment algorithm for a multi-period transportation problem. Same J. 16 517-524.

[14.59] Briggs, F. E. A. (1962). Solution of the Hitchcock problem with one single row capacity constraint per row by the Ford-Fulkerson method. Same J. 9 107-120.

[14.60] Charnes, A., F. Glover, and D. Klingman (1971).

The lower bounded and partial upper bounded distribution model. Same J. 18 277-281.

[14.61] Ford, L. R., Jr., and D. R. Fulkerson (1957). A primal dual algorithm for the capacitated Hitchcock problem. Same J. 4 47-54.

[14.62] Gaddum, J. W., A. J. Hoffman, and D. Sokolowsky (1954). On the solution of the caterer problem. Same J. 1 223-229.

[14.63] Galler, B. A., and P. S. Dwyer (1957). Translating the method of reduced matrices to machines. Same J. 4 55-71.

[14.64] Garfinkel, R. S., and M. R. Rao (1971). The bottleneck transportation problem. Same J. 18 465-472.

[14.65] Gassner, B. J. (1964). Cycling in the Transportation problem. Same J. 11 43-58.

[14.66] Glicksman, S., L. Johnson, and L. Eselson (1960). Coding the transportation problem. Same J. 7 169-183.

[14.67] Hammer, P. L. (1969). Time minimizing transportation problems. Same J. 16 345-357.

[14.68] Hammer, P. L. (1971). Communication on "the bottleneck transportation problem" and "some remarks on the time transportation problem." Same J. 18 487-490.

[14.69] Hoffman, A. J., and H. M. Markowitz (1963). A note on shortest path, assignment, and transportation problems. Same J. 10 375-379.

[14.70] Holladay, J. (1964). Some transportation problems and techniques for solving them. Same J. 11 15-42.

[14.71] Jacobs, W. (1954). The caterer problem. Same J. 1 154-165.

[14.72] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. Same J. 2 83-97.

[14.73] Kuhn, H. W. (1956). Variants of the Hungarian method for assignment problems. Same J. 3 253-258.

[14.74] Lagemann, J. J. (1967). A method for solving the transportation problem. Same J. 14 89-99.

[14.75] Prager, W. (1957). Numerical solution of the generalized transportation problem. Same J. 4 253-261.

[14.76] Rigby, F. D. (1962). An analog and derived algorithm for the dual transportation problem. Same J. 9 81-96.

[14.77] Simonnard, M. A., and G. F. Hadley (1959). Maximum number of iterations in the transportation problem. Same J. 6 125-129.

[14.78] Szwarc, W. (1971). The transportation paradox. Same J. 18 185-202.

[14.79] Szwarc, W. (1971). Some remarks on the time transportation problem. Same J. 18 473-485.

Integer Programming. An important special case of mathematical programming requires that elements, or all, of the solution be integer valued. Many problems to be modeled require allocation of indivisible units, such as men, ships, or aircraft, or include decision variables that can be represented as (0,1) integer variables. Although distribution and assignment problems in integers produce integer solutions, this situation does not hold for more general cases. As it is known that continuous approximations to discrete problems can be distinctly suboptimal, exact integer solutions become of interest. R. E. Gomory's cutting plane algorithm and solution of the integer programming problem of 1958 was a major development in the theory of integer linear programming and most subsequent work springs from this development. Some examples will be

found below. Articles in the Quarterly have dealt with alternative cutting plane approaches, solution characterization, and the central question of finite convergence. An example is Gomory and Hoffman's [14.82] which characterizes Dantzig cuts and shows the process not to converge in the general case. Bowman and Nemhauser in [14.80] show that Charnes and Cooper's modified cuts do give convergence.

[14.80] Bowman, V. J., Jr., and G. L. Nemhauser (1970). A finiteness proof for modified Dantzig cuts in integer programming. Naval Res. Logist. Quart. 17 309-313.

[14.81] Dantzig, G. B. (1959). Note on solving linear programs in integers. Same J. 6 75-76.

[14.82] Gomory, R. E., and A. J. Hoffman (1963). On the convergence of an integer programming process. Same J. 10 121-123.

[14.83] Robillard, P. (1971). (0,1) hyperbolic programming problems. Same J. 18 47-57.

[14.84] Salkin, H. M., and P. Breining (1971). Integer points on the Gomory fractional cut (hyperplane). Same J. 18 491-496.

Economics. Problems of economics have been closely associated with the development of mathematical programming. Linear programming economic models and economic interpretations in mathematical programming were developed early in the history of linear programming. Leontief's development of input-output models of interindustry economics (1936) was an important related factor that served as one of the spurs to the development of linear programming. Input-output models were subsequently fused with mathematical programming models to produce powerful econometric models.

Articles in the Quarterly have dealt with Leontief models, linear and goal programming applications in economics, and characterization and extension of von Neumann's economic equilibrium model. The von Neumann model of 1937 is of current interest for its treatment

of an expanding, rather than static, economy and its provision of growth index possibly superior to GNP (see Morgenstern and Thompson [14.90]). Another important article was Gale's [14.88], which established a class of examples for which the competitive equilibrium is not globally stable.

[14.85] Enzer, H., S. D. Berry, and J. I. Martin, Jr. (1968). Economic impact and the notion of compensated procurement. Naval Res. Logist. Quart. 15 63-79.

[14.86] Enzer, H., and D. C. Dellinger (1968). On some economic concepts of multiple incentive contracting. Same J. 15 477-489.

[14.87] Frisch, H. (1969). Consumption, the rate of interest and the rate of growth in the von Neumann model. Same J. 16 459-484.

[14.88] Gale, D. (1963). A note on global instability of competitive equilibrium. Same J. 10 81-87.

[14.89] Manne, A. S. (1960). Comments on "inter-industry economics" by Chenery and Clark. Same J. 7 385-389.

[14.90] Morgenstern, O., and G. L. Thompson (1969). An open expanding economy model. Same J. 16 443-457.

[14.91] Quandt, R. E. (1958). Probabilistic errors in the Leontief system. Same J. 5 155-170.

[14.92] Quandt, R. E. (1959). On the solution of probabilistic Leontief systems. Same J. 6 295-305.

[14.93] Spivey, W. A., and H. Tamura (1970). Goal programming in econometrics. Same J. 17 183-192.

[14.94] Wong, Y. K. (1954). An elementary treatment of an input-output system. Same J. 1 321-326.

[14.95] Wurtele, Z. S. (1961). A note on pyramid building. Same J. 8 377-379.

Quadratic and Quadratic Assignment Problems. Quadratic programming, optimization of a second degree objective function subject to linear constraints, was a natural extension of linear programming. The quadratic objective function is attractive as a second-order approximation of more general functions and has a physical basis in terms of euclidean distance, area, power, and so on. Specialized algorithms were developed for this case because of its importance and relative accessibility. Articles in the Quarterly have dealt with algorithms and solution approaches to this problem and the special case of the quadratic assignment problem. An important early algorithm was given by M. Frank and P. Wolfe in [14.97]. Their algorithm solves a linear program at each iteration to determine step size along the gradient from the current point, relying for successful implementation upon efficient LP codes. Because of slow or nonconvergence numerically sometimes, other algorithms by Beale, van de Panne, Whinston, and others, have come to the fore.

[14.96] Beale, E. M. L. (1959). On quadratic programming. Naval Res. Logist. Quart. 6 227-243.

[14.97] Frank, M., and P. Wolfe (1956). An algorithm for quadratic programming. Same J. 3 95-110.

[14.98] Gaschutz, G. K., and J. H. Ahrens (1968). Suboptimal algorithms for the quadratic assignment problem. Same J. 15 49-62.

[14.99] Greenberg, H. (1969). A quadratic assignment problem without column constraints. Same J. 16 417-421.

[14.100] Hildreth, C. (1957). A quadratic programming procedure. Same J. 4 79-85.

[14.101] Markowitz, H. (1956). The optimization of a quadratic function subject to linear constraints. Same J. 3 111-133.

[14.102] Pierce, J. F., and W. B. Crowston (1971). Tree-search algorithms for quadratic assignment problems.

Same J. 18 1-36.

[14.103] van de Panne, C., and A. Whinston (1964). Simplicial methods for quadratic programming. Same J. 11 273-302.

[14.104] Whinston, A. (1965). The bounded variable problem--an application of the dual method for quadratic programming. Same J. 12 173-180.

Probabilistic Programming. The problems we have been been examining can be called deterministic in that parameters are assumed to be known constants. In many situations parameters are either not known precisely or are subject to variation as, for example, future demand for goods or services. There have been several approaches developed to deal with these problems. Among these are G. B. Dantzig's "linear programming under uncertainty" of 1955 and A. Charnes, W. W. Cooper, and G. H. Symond's "chance constrained programming" of 1954. Examples of both approaches will be found in the articles listed below. One article of importance here is Charnes and Cooper's [14.107] which further developed the theory of chance constrained programming in the context of investigating forward planning problems for time-phased transport requirements.

[14.105] Baron, D. P. (1971). Information in two-stage programming under uncertainty. Naval Res. Logist. Quart. 18 169-176.

[14.106] Bracken, J., and R. M. Soland (1966). Statistical decision analysis of stochastic linear programming problems. Same J. 13 205-225.

[14.107] Charnes, A., and W. W. Cooper (1960). Some problems and models for time-phased transport requirements, chance constrained programs with normal deviates and linear decision rules. Same J. 7 533-544.

[14.108] Midler, J. L., and R. D. Wollmer (1969). Stochastic programming models for scheduling airlift operations. Same J. 16 315-330.

[14.109] Yechiali, U. (1971). A note on a stochastic production-maximizing transportation problem. Same J. 18 429-431.

Mathematical Programming Results. The Quarterly has served as a forum for a variety of articles describing theoretical results in mathematical programming. Topics have included constraint qualification, duality, and equivalent formulations in nonlinear and convex programming, semi-infinite programming, linear fractional programming, linear programming equivalences, and dynamic programming. A major article is Karlin's [14.125] but, the work referenced below of Arrow, Hurwicz and Uzawa, Bellman, Charnes and Cooper, and Shapley indicate the great scope and variety included.

[14.110] Arrow, K. J., L. Hurwicz, and H. Uzawa (1961). Constraint qualifications in maximization problems. Naval Res. Logist. Quart. 8 175-191.

[14.111] Bellman, R. (1956). Notes on the theory of dynamic programming IV--maximization over discrete sets. Same J. 3 67-70.

[14.112] Bellman, R. (1960). Functional equations and successive approximations in linear and nonlinear programming. Same J. 7 63-83.

[14.113] Charnes, A., and W. W. Cooper (1958). Nonlinear network flows and convex programming over incidence matrices. Same J. 5 231-240.

[14.114] Charnes, A., and W. W. Cooper (1962). Programming with linear fractional functionals. Same J. 9 181-186.

[14.115] Charnes, A., and W. W. Cooper (1968). Structural sensitivity analysis in linear programming and an exact product form left inverse. Same J. 15 517-522.

[14.116] Charnes, A., W. W. Cooper, and K. O. Kortanek (1969). On the theory of semi-infinite programming and a generalization of the Kuhn-Tucker saddle point theorem

for arbitrary convex functions. Same J. 16 41-51.

[14.117] Charnes, A., W. W. Cooper, and M. Miller (1961). Dyadic programs and subdual methods. Same J. 8 1-23.

[14.118] Charnes, A., and C. E. Lemke (1954). Minimization of non-linear separable convex functionals. Same J. 1 301-312.

[14.119] D'Esopo, D. A. (1959). A convex programming procedure. Same J. 6 33-42.

[14.120] Ericson, W. A. (1968). On the minimization of a certain convex function arising in applied decision theory. Same J. 15 33-48.

[14.121] Evans, J. P. (1970). On constraint qualifications in nonlinear programming. Same J. 17 281-286.

[14.122] Gale, D. (1956). The basic theorems of real linear equations, inequalities, linear programming and game theory. Same J. 3 193-200.

[14.123] Hoffman, A. J. (1963). On abstract dual linear programs. Same J. 10 369-373.

[14.124] Joksch, H. C. (1964). Programming with fractional linear objective functions. Same J. 11 197-204.

[14.125] Karlin, S. (1955). The structure of dynamic programming models. Same J. 2 285-294.

[14.126] Karush, W. (1959). A theorem in convex programming. Same J. 6 245-260.

[14.127] Malik, H. J. (1968). A note on generalized inverses. Same J. 15 605-612.

[14.128] Mangasarian, O. L. (1963). Equivalence in nonlinear programming. Same J. 10 299-306.

[14.129] Martos, B. (Translated by A. and V. Whinston)

(1964). Hyperbolic programming. Same J. 11 135-155.

[14.130] Randolph, P. H., and G. E. Swinson (1969). The discrete max-min problem. Same J. 16 309-314.

[14.131] Ritter, K., (Translated by M. Meyer) (1967). A method for solving nonlinear maximum problems depending on parameters. Same J. 14 147-162.

[14.132] Saaty, T. L. (1968). On nonlinear optimization in integers. Same J. 15 1-22.

[14.133] Shapley, L. S. (1962). Complements and substitutes in the optimal assignment problem. Same J. 9 45-48.

[14.134] Taylor, R. J., and S. P. Thompson (1958). On a certain problem in linear programming. Same J. 5 171-187.

[14.135] Whinston, A. (1965). Conjugate functions and dual programs. Same J. 12 315-322.

[14.136] Zionts, S. (1968). Programming with linear fractional functionals. Same J. 15 449-452.

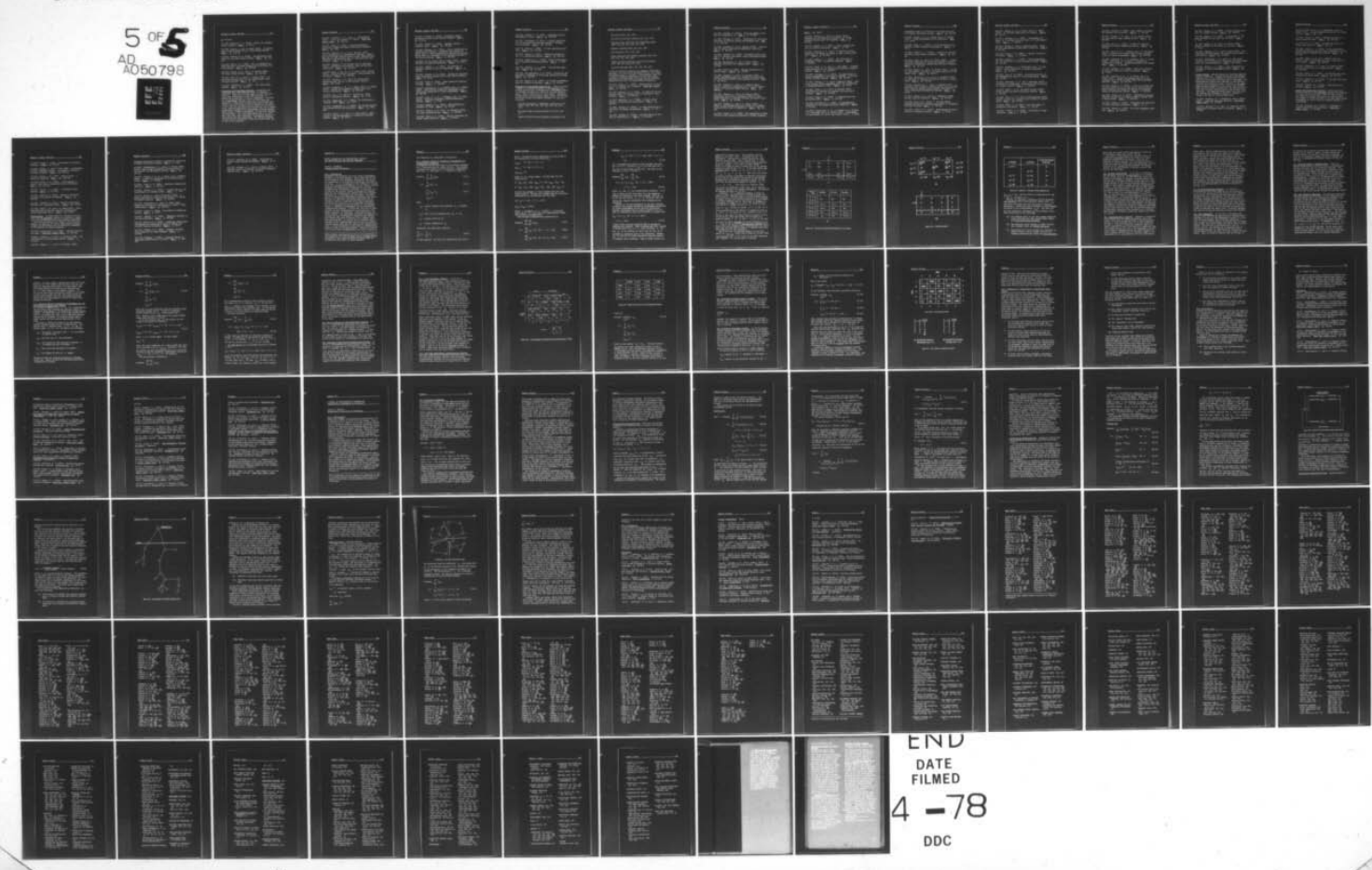
[14.137] Zwart, P. B. (1970). Nonlinear programming--the choice of direction by gradient projection. Same J. 17 431-438.

Fixed Charge Problems. Fixed charge problems are characterized by a fixed cost that is incurred any time an activity is operated at a nonzero level, in addition to variable costs. Minimum plant investments and set-up charges are examples of this situation. Articles in the Quarterly have dealt with formulations of the fixed charge and fixed charge transportation problems, and with algorithms for exact and approximate solutions. Hirsch and Dantzig's [14.143] is a publication of a 1954 Rand Corporation paper that was among the first treatments of this problem and is of historical interest.

[14.138] Almogy, Y., and O. Levin (1971). The fractional fixed charge problem. Naval Res. Logist. Quart.

AD-A050 798 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 15/5
MODERN TRENDS IN LOGISTICS RESEARCH. PROCEEDINGS OF A CONFERENC--ETC(U)
1976 W H MARLOW N00014-75-C-0729
UNCLASSIFIED NL

5 OF 5
AD
AD050798



END
DATE FILMED
4 -78
DDC

18 307-315.

[14.139] Balinski, M. L. (1961). Fixed cost transportation problems. Same J. 8 41-54.

[14.140] Cooper, L., and C. Drebes (1967). An approximate solution method for the fixed charge problem. Same J. 14 101-113.

[14.141] Denzler, D. R. (1969). An approximate algorithm for the fixed charge problem. Same J. 16 411-416.

[14.142] Dwyer, P. S. (1966). Use of completely reduced matrices in solving transportation problems with fixed charges. Same J. 13 289-313.

[14.143] Hirsch, W. M., and G. B. Dantzig (1968). The fixed charge problem. Same J. 15 413-424.

[14.144] Kuhn, H. W., and W. J. Baumol (1962). An approximative algorithm for the fixed-charges transportation problem. Same J. 9 1-15.

[14.145] Steinberg, D. I. (1970). The fixed charge problem. Same J. 17 217-235.

Sequencing and Scheduling Problems. Sequencing and scheduling have been important areas of logistics research. These cover a wide range of practical problems, from scheduling of individual machines, to production scheduling, to project management. Articles in the Quarterly indicate a part of this range and diversity. These articles have dealt with problem formulation and extensions, solution approaches, and algorithms, and have included several reviews of the literature. Solution approaches here include the use of linear programming, probabilistic programming, network methods, graph theory, dynamic programming, critical path methods, and simulation. The first major paper was S. M. Johnson's [14.162], which presented the original results on multi-stage job shop scheduling. The Bartlett-Charnes [14.148] cyclic scheduling results are still unique and in use in the railroads.

[14.146] Arthanari, T. S., and A. C. Mukhopadhyay (1971). A note on a paper by W. Szwarc. Naval Res. Logist. Quart. 18 135-138.

[14.147] Balas, E. (1970). Machine sequencing: disjunctive graphs and degree-constrained subgraphs. Same J. 17 1-10.

[14.148] Bartlett, T. E., and A. Charnes (1957). Cyclic scheduling and combinational topology: assignment and routing of motive power to meet scheduling and maintenance requirements; Part II, generalization and analysis. Same J. 4 207-220.

[14.149] Bratley, P., M. Florian, and P. Robillard (1971). Scheduling with earliest start and due date constraints. Same J. 18 511-519.

[14.150] Burt, J. M., Jr., D. P. Gaver, and M. Perlas (1970). Simple stochastic networks: some problems and procedures. Same J. 17 439-459.

[14.151] Buzacott, J. A., and S. K. Dutta (1971). Sequencing many jobs on a multi-purpose facility. Same J. 18 75-82.

[14.152] Cremeans, J. E., R. A. Smith, and G. R. Tyndall (1970). Optimal multicommodity network flows with resource allocation. Same J. 17 269-279.

[14.153] Day, J. E., and M. P. Hottenstein (1970). Review of sequencing research. Same J. 17 11-39.

[14.154] Elmaghraby, S. E. (1968a). The sequencing of "related" jobs. Same J. 15 23-32.

[14.155] Elmaghraby, S. E. (1968b). The machine sequencing problem--review and extensions. Same J. 15 205-232.

[14.156] Evans, J. P., and F. J. Gould (1971). Application of the GLM technique to a production planning problem. Same J. 18 59-74.

[14.157] Giffler, B. (1963). Scheduling general production systems using schedule algebra. Same J. 10 237-255.

[14.158] Giffler, B. (1968). Schedule algebra: a progress report. Same J. 15 255-280.

[14.159] Gleaves, V. B. (1957). Cyclic scheduling and combinational topology: assignment and routing of motive power to meet scheduling and maintenance requirements; Part I, a statement of the operation problem of the Frisco Railroad. Same J. 4 203-205.

[14.160] Hu, Te Chiang, and W. Prager (1959). Network analysis of production smoothing. Same J. 6 17-23.

[14.161] Jackson, J. R. (1956). An extension of Johnson's results on job lot scheduling. Same J. 3 201-204.

[14.162] Johnson, S. M. (1954). Optimal two and three stage production schedules with setup times included. Same J. 1 61-68.

[14.163] Klein, M. (1957). Some production planning problems. Same J. 4 269-286.

[14.164] Kortanek, K. O., D. Sodaro, and A. L. Soyster (1968). Multi-product production scheduling via extreme point properties of linear programming. Same J. 15 287-300.

[14.165] Levy, F. K., G. L. Thompson, and J. D. Wiest (1962). Multiship, multishop, workload-smoothing program. Same J. 9 37-44.

[14.166] Maxwell, W. L. (1964). The scheduling of economic lot sizes. Same J. 11 89-124.

[14.167] O'Neill, R. R., and J. K. Weinstock (1954). The cargo handling system. Same J. 1 282-288.

[14.168] Smith, W. E. (1956). Various optimizers for single stage production. Same J. 3 59-66.

[14.169] Spinner, A. H. (1968). Sequencing theory--development to date. Same J. 15 319-330.

[14.170] Srinivasan, V. (1971). A hybrid algorithm for the one machine sequencing problem to minimize total tardiness. Same J. 18 317-327.

[14.171] Szwarc, W. (1968). On some sequencing problems. Same J. 15 127-156.

[14.172] Szwarc, W. (1971). Elimination methods in the m x n sequencing problem. Same J. 18 295-305.

[14.173] Thompson, G. L. (1960). Recent developments in the job-shop scheduling problem. Same J. 7 585-589.

[14.174] Thompson, G. L. (1968). CPM and DCPM under risk. Same J. 15 233-240.

[14.175] von Lanzenauer, C. H. (1970). Production and employment scheduling in multistage production systems. Same J. 17 193-198.

[14.176] Wagner, H. M. (1959). An integer programming model for machine scheduling. Same J. 6 131-140.

Mathematical Programming Applications. There have been a number of articles published in the Quarterly describing applications of mathematical programming models to problems in logistics, viewed in its most general sense. Some of these will be found below. Categories of application have included, among others, the following where we omit the Chapter prefixes 14. from the citations.

Aircraft procurement, deployment, operations, and rework (188, 189, 190, 198, 200, 203, 207, 229, 232)

The Naval tanker routing problem (178, 185, 195, 217)

Commercial and military shipping requirements and

allocation (199, 212, 219)

The Naval electronics problem (222, 233, 235)

Transportation allocation and scheduling models (186, 193, 225, 227, 230, 234, 236)

Manpower planning (205, 216, 221, 228)

Bid evaluation (179, 204, 240)

Other financial and economic problems (184, 196, 209, 211, 231, 238, 239)

Supply system and resource allocation problems (177, 187, 206, 208, 213, 215)

Location problems (192, 197, 220, 223, 241)

A typical article of interest is Laderman, Gleiberman, and Egan's [14.219] which describes the use of a linear programming model to schedule a heterogeneous fleet for a season's shipping requirements on the Great Lakes. Experience with this model in actual use is discussed. Bartlett's [14.178] is of pathbreaking originality.

[14.177] Allen, S. G. (1958). Redistribution of total stock over several user locations. Naval Res. Logist. Quart. 5 337-345.

[14.178] Bartlett, T. E. (1957). An algorithm for the minimum number of transport units to maintain a fixed schedule. Same J. 4 139-149.

[14.179] Beged-Dov, A. G. (1970). Contract award analysis by mathematical programming. Same J. 17 297-307.

[14.180] Bellman, R. (1954a). On some applications of the theory of dynamic programming. Same J. 1 141-153.

[14.181] Bellman, R. (1954b). Decision making in the face of uncertainty--I. Same J. 1 230-232.

[14.182] Bellman, R. (1954c). Decision making in the face of uncertainty--II. Same J. 1 327-332.

[14.183] Bellman, R. (1957). Formulation of recurrence equations for shuttle process and assembly line. Same J. 4 321-334.

[14.184] Bellman, R., and S. Dreyfus (1958). A bottleneck situation involving interdependent industries. Same J. 5 307-314.

[14.185] Bellmore, M. (1968). A maximum utility solution to a vehicle constrained tanker scheduling problem. Same J. 15 403-412.

[14.186] Bennington, G., and S. Lubore (1970). Resource allocation for transportation. Same J. 17 471-484.

[14.187] Blitz, M. (1963). Optimum allocation of a spares budget. Same J. 10 175-191.

[14.188] Bracken, J., and P. R. Burnham (1968). A linear programming model for minimum-cost procurement and operation of Marine Corps training aircraft. Same J. 15 81-97.

[14.189] Bracken, J., and J. D. Longhill (1964). Note on a model for minimizing cost of aerial tankers support of a practice bomber mission. Same J. 11 359-364.

[14.190] Bracken, J., and K. W. Simmons (1966). Minimizing reductions in readiness caused by time-phased decreases in aircraft overhaul and repair activities. Same J. 13 159-165.

[14.191] Bracken, J., and T. C. Varley (1963). A model for determining protection levels for equipment classes within a set of subsystems. Same J. 10 257-262.

[14.192] Breuer, M. A. (1966). The formulation of some allocation and connection problems as integer programs.

Same J. 13 83-95.

[14.193] Charnes, A., and M. H. Miller (1957). Mathematical programming and evaluation of freight shipment systems, part II--analysis. Same J. 4 243-252.

[14.194] Cheney, L. K. (1957). Linear program planning of refinery operations. Same J. 4 9-16.

[14.195] Dantzig, G. B., and D. R. Fulkerson (1954). Minimizing the number of tankers to meet a fixed schedule . Same J. 1 217-222.

[14.196] Daubin, S. C. (1958). The allocation of development funds: an analytic approach. Same J. 5 263-276.

[14.197] Davis, P. S., and T. L. Ray (1969). A branch-bound algorithm for the capacitated facilities location problem. Same J. 16 331-344.

[14.198] Dellinger, D. C. (1971). An application of linear programming to contingency planning: a tactical airlift systems analysis. Same J. 18 357-378.

[14.199] D'Esopo, D. A., and B. Lefkowitz (1964). Note on an integer linear programming model for determining a minimum embarkation fleet. Same J. 11 79-82.

[14.200] Donis, J. N., and S. M. Pollock (1967). Allocation of resources to randomly occurring opportunities. Same J. 14 513-527.

[14.201] Evans, G. W. (1958). A transportation and production model. Same J. 5 137-154.

[14.202] Firstman, S. I. (1960). An approximation algorithm for an optimum aim-points algorithm. Same J. 7 151-167.

[14.203] Fitzpatrick, G. R., J. Bracken, M. J. O'Brien, L. G. Wentling, and J. C. Whiton (1967). Programming the procurement of airlift and sealift forces: a linear

programming model for analysis of the least-cost mix of strategic deployment systems. Same J. 14 241-255.

[14.204] Gainen, L., D. P. Honig, and E. D. Stanley (1954). Linear programming in bid evaluation. Same J. 1 49-54.

[14.205] Gass, S. I. (1957). On the distribution of manhours to meet scheduled requirements. Same J. 4 17-25.

[14.206] Gilbert, J. C. (1964). A method of resource allocation using demand preference. Same J. 11 217-225.

[14.207] Gross, D., and R. M. Soland (1969). A branch and bound algorithm for allocation problems in which constraint coefficients depend upon decision variables. Same J. 16 157-174.

[14.208] Hadley, G., and T. M. Whitin (1961). A model for procurement, allocation, and redistribution for low demand items. Same J. 8 395-414.

[14.209] Hitchcock, D. F., and J. B. MacQueen (1970). On computing the expected discounted return in a Markov chain. Same J. 17 237-241.

[14.210] Houston, B. F., and R. A. Huffman (1971). A technique which combines modified pattern search methods with composite designs and polynomial constraints to solve constrained optimization problems. Same J. 18 91-98.

[14.211] Howard, G. T., and G. L. Nemhauser (1968). Optimal capacity expansion. Same J. 15 535-550.

[14.212] Hunt, R. B., and E. F. Rosholdt (1960). Determining merchant shipping requirements in integrated military planning. Same J. 7 545-575.

[14.213] Jewell, W. S. (1957). Warehousing and distribution of a seasonal product. Same J. 4 29-34.

[14.214] Karlin, S., W. E. Pruitt, and W. G. Madow (1963). On choosing combinations of weapons. Same J. 10 95-119.

[14.215] Karreman, H. F. (1960). Programming the supply of a strategic material--part I, a nonstochastic model. Same J. 7 261-279.

[14.216] Karush, W., and A. Vazsonyi (1957). Mathematical programming and employment scheduling. Same J. 4 297-320.

[14.217] Kelley, J. E. (1955). A dynamic transportation model. Same J. 2 175-180.

[14.218] Kolesar, P. J. (1967). Linear programming and the reliability of multicomponent systems. Same J. 14 317-327.

[14.219] Laderman, J., L. Gleiberman, and J. F. Egan (1966). Vessel allocation by linear programming. Same J. 13 315-320.

[14.220] Love, R. F. (1969). Locating facilities in three-dimensional space by convex programming. Same J. 16 503-516.

[14.221] McCloskey, J. F., and F. Hanssmann (1957). An analysis of stewardess requirements and scheduling for a major airline. Same J. 4 183-191.

[14.222] McShane, R. E., and Henry Solomon (1958). Letter concerning Naval electronics problem. Same J. 5 363-367.

[14.223] Nair, K. P. K., and R. Chandrasekaran (1971). Optimal location of a single service center of certain types. Same J. 18 503-510.

[14.224] Noble, S. B. (1960). Some flow models of production constraints. Same J. 7 401-419.

[14.225] O'Neill, R. R. (1960). Scheduling of cargo containers. Same J. 7 577-584.

[14.226] Pollack, M. (1958). Some studies on shuttle and assembly-line processes. Same J. 5 125-136.

[14.227] Pruzan, P. M., and J. T. R. Jackson (1967). The many-product cargo loading problem. Same J. 14 381-390.

[14.228] Rau, J. G. (1971). A model for manpower productivity during organization growth. Same J. 18 543-559.

[14.229] Rice, E. W., J. Bracken, and A. W. Pennington (1971). Allocation of carrier-based attack aircraft using non-linear programming. Same J. 18 379-393.

[14.230] Saposnik, R., A. R. Lindeman, and V. L. Smith (1959). Allocation of a resource to alternative probabilistic demands: transport-equipment pool assignments. Same J. 6 193-207.

[14.231] Scherer, F. M. (1966). Time-cost tradeoffs in uncertain empirical research projects. Same J. 13 71-82.

[14.232] Schwartz, A. N., J. A. Sheler, and C. R. Cooper (1971). Dynamic programming approach to the optimization of naval aircraft rework and replacement policies. Same J. 18 395-414.

[14.233] Smith, J. W. (1956). A plan to allocate and procure electronic sets by the use of linear programming techniques and analytical methods of assigning values to qualitative factors. Same J. 3 151-162.

[14.234] Soyster, H. R. (1957). Mathematical programming and evaluations of freight shipment systems, part I--applications. Same J. 4 237-242.

[14.235] Suzuki, G. (1957). Procurement and allocation of naval electronic equipments. Same J. 4 1-7.

[14.236] Szwarc, W. (1967). The truck assignment problem. Same J. 14 529-557.

[14.237] Vergin, R. C. (1968). Optimal renewal policies for complex systems. Same J. 15 523-534.

[14.238] Wagner, H. M. (1960). A postscript to "dynamic problems in the theory of the firm." Same J. 7 7-12.

[14.239] Wagner, H. M., and T. M. Whitin (1958). Dynamic problems in the theory of the firm. Same J. 5 53-74.

[14.240] Waggener, H. A., and G. Suzuki (1967). Bid evaluation for procurement of aviation fuel at DFSC: a case history. Same J. 14 115-129.

[14.241] Wesolowsky, G. O., and R. F. Love (1971). Location of facilities with rectangular distances among point and area destinations. Same J. 18 83-90.

[14.242] Whiton, J. C. (1967). Some constraints on shipping in linear programming models. Same J. 14 257-260.

Network Theory. Network theory has important applications in logistics. Transportation and transshipment problems are but two examples of cases where network models and techniques have been used to advantage. The Ford-Fulkerson network flow approach to distribution problems was previously noted. Articles in the Quarterly include treatments of the convex cost minimum flow network problem, interdiction and isolation problems with military interpretations, and network flow functions. One important early article in this field was Dantzig and Fulkerson's [14.244]. Shapley's incisive [14.252] is also included below.

[14.243] Bellmore, M., G. Bennington, and S. Lubore (1970). A network isolation algorithm. Naval Res. Logist. Quart. 17 461-469.

[14.244] Dantzig, G. B., and D. R. Fulkerson (1955). Computation of maximal flows in networks. Same J. 2 277-283.

[14.245] Ghare, P. M., D. C. Montgomery, and W. C. Turner (1971). Optimal interdiction policy for a flow network. Same J. 18 37-45.

[14.246] Glover, F. (1967). Maximum matching in a convex bipartite graph. Same J. 14 313-316.

[14.247] Hu, T. C. (1966). Minimum-cost flows in convex-cost networks. Same J. 13 1-9.

[14.248] Jarvis, J. J. (1969). On the equivalence between the node-arc and arc-chain formulations for the multi-commodity maximal flow problem. Same J. 16 525-529.

[14.249] Lubore, S. H., H. D. Ratliff, and G. T. Sicilia (1971). Determining the most vital link in a flow network. Same J. 18 497-502.

[14.250] McMasters, A. W., and T. M. Mustin (1970). Optimal interdiction of a supply network. Same J. 17 261-268.

[14.251] Menon, V. V. (1965). The minimal cost flow problem with convex costs. Same J. 12 163-172.

[14.252] Shapley, L. S. (1961). On network flow functions. Same J. 8 151-158.

[14.253] Wollmer, R. D. (1970). Interception in a network. Same J. 17 207-216.

Search and Surveillance Problems. There have been several articles in the Quarterly dealing with problems of search and surveillance. Some of those articles that employed extremal principles are listed below. One example is Derman and Klein's [14.256] where linear programming is used to find the optimal sequence of inspections.

[14.254] Blachman, N. M. (1959). Prolegomena to optimum discrete search procedures. Naval Res. Logist. Quart. 6 273-281.

[14.255] Derman, C. (1961). On minimax surveillance schedules. Same J. 8 415-419.

[14.256] Derman, C., and M. Klein (1966). Surveillance of multi-component systems: a stochastic travelling salesman's problem. Same J. 13 103-111.

[14.257] Dobbie, J. M. (1963). Search theory: a sequential approach. Same J. 10 323-334.

[14.258] Enslow, P. H. (1966). A bibliography of search theory and reconnaissance theory literature. Same J. 13 177-202.

[14.259] Isbell, J. R. (1957). An optimal search pattern. Same J. 4 357-359.

[14.260] Roeloffs, R. (1963). Minimax surveillance schedules with partial information. Same J. 10 307-322.

[14.261] Roeloffs, R. (1967). Minimax surveillance schedules for replaceable units. Same J. 14 461-471.

[14.262] Smith, M. W., and J. E. Walsh (1971). Optimum sequential search with discrete locations and random acceptance ratios. Same J. 18 159-167.

Miscellaneous. Included below is a selection of articles from the Quarterly related to a variety of logistical or mathematical issues not elsewhere classified.

An article of interest is O'Neill's [14.275] which presents an analysis of cyclic linked cargo handling systems and demonstrates use of simulation techniques in the solution of several examples.

[14.263] Antosiewicz, H. A. (1955). Analytic study of war games. Naval Res. Logist. Quart. 2 181-208.

[14.264] Aumann, R. J., and J. B. Kruskal (1958). The coefficients in an allocation problem. Same J. 5 111-123.

[14.265] Aumann, R. J., and J. B. Kruskal (1959).

Assigning quantitative values to qualitative factors in the Naval electronics problem. Same J. 6 1-16.

[14.266] Brandenburg, R. G., and A. C. Stedry (1966). Toward a multi-stage information conversion model of the research and development process. Same J. 13 129-146.

[14.267] D'Esopo, D. A., H. L. Dixon, and B. Lefkowitz (1960). A model for simulating an air-transportation system. Same J. 7 213-220.

[14.268] Flood, M. M. (1958). Operations research and logistics. Same J. 5 323-335.

[14.269] Gourary, M. H. (1958). A simple rule for the consolidation of allowance lists. Same J. 5 1-15.

[14.270] Harary, F., and M. Richardson (1959). A matrix algorithm for solutions and r-bases of a finite irreflexive relation. Same J. 6 307-314.

[14.271] Hershkowitz, M., and S. B. Noble (1965). Finding the inverse and connections of a type of large sparse matrix. Same J. 12 119-132.

[14.272] Isaacs, R. (1955). The problem of aiming and evasion. Same J. 2 47-67.

[14.273] Jackson, J. R. (1957). Simulation research on job shop production. Same J. 4 287-295.

[14.274] Mellon, W. G. (1958). A selected, descriptive bibliography of references on priority systems and related, nonprice allocators. Same J. 5 17-27.

[14.275] O'Neill, R. R. (1957). Analysis and Monte Carlo simulation of cargo handling. Same J. 4 223-236.

[14.276] Pfanzagl, J. (1959). A general theory of measurement-applications to utility. Same J. 6 283-294.

[14.277] Salveson, M. E. (1961). Principles of dynamic weapon systems programming. Same J. 8 79-110.

[14.278] Wortham, A. W., and E. B. Wilson (1971). A backward recursive technique for optimal sequential sampling plans. Same J. 18 203-213.

Chapter 15

RECENT THEORETICAL AND COMPUTATIONAL RESULTS FOR TRANSPORTATION AND RELATED PROBLEMS*

Gerald L. Thompson
Carnegie-Mellon University

15.1 Introduction

Although it was probably the first linear programming problem to be stated formally, and in many ways is one of the simplest such problems, research on the theory and computation of transportation problems continues to be active. It is the purpose of this chapter to discuss these and analogous developments for two closely related problems--the generalized transportation and time transportation (bottleneck) problems. In Section 15.2 we discuss the operator theory of parametric programming for the transportation problem which was recently developed by V. Srinivasan and the author. The dual matrix is defined and managerial interpretations given. Rim and cost operators are defined and their use in solving a variety of applications briefly discussed. In Section 15.3 we carry out an analogous discussion due to V. Balachandran and the author of the operator theory of parametric programming for the generalized transportation problem. In this case the idea of basic solution is more complicated and a new operator, the weight operator, is developed. Some applications are also included. The algorithms of Szwarc and Hammer for solving the time (bottleneck) transportation problem are briefly discussed in Section 15.4, together with some applications. In Section 15.5 we present some of the most recent computational results with the primal transportation codes and the Szwarc time transportation code. A brief discussion is also made of computational results of one of the dual methods. Finally in Section 15.6 we give some recommendations for future studies on

*This chapter was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract N00014-67-A09314-0007 NR 047-048 with the Office of Naval Research.

the comparison of algorithmic efficiencies.

15.2 Operator Theory of Parametric Programming for Transportation Problems

We summarize briefly here some of the results obtained by V. Srinivasan and the author [15.27]. The standard transportation problem for distributing a homogeneous good from m warehouses to n markets is given by

$$\text{minimize} \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{15.1}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = a_i \tag{15.2}$$

$$\sum_{i=1}^{n} x_{ij} = b_j \tag{15.3}$$

$$x_{ij} \geq 0$$

where

x_{ij} = amount shipped from warehouse W_i to market M_j

c_{ij} = unit cost of shipping from W_i to M_j

a_i = amount stored in W_i

b_j = amount demanded in M_j

Frequently the additional condition

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \tag{15.4}$$

is also imposed. We shall not assume here that (15.4)

holds. Instead we add an additional row and column to the problem with the cost definitions

$$c_{i,n+1} = 0 \quad \text{for} \quad i = 1,\ldots,m$$

$$c_{m+1,j} = 0 \quad \text{for} \quad j = 1,\ldots,n$$

$$c_{m+1,n+1} = M$$

where M is a large number. We also make the rim conditions be

$$\text{If} \quad \sum a_i \geq \sum b_j \quad \text{then} \quad a_{m+1} = 0 \quad \text{and} \quad b_{n+1} = \sum a_i - \sum b_j$$

$$\text{If} \quad \sum a_i < \sum b_j \quad \text{then} \quad a_{m+1} = \sum b_j - \sum a_i \quad \text{and} \quad b_{n+1} = 0$$

so that the analogue to (15.4) always holds for the larger problem. We also will assume that the following perturbation of coefficients has been made:

$$a_i' = a_i + \varepsilon \quad \text{for} \quad i = 1,\ldots,m+1$$

$$b_{n+1}' = b_{n+1} + (m+1)\varepsilon$$

where ε satisfies $0 < \varepsilon < 1/(m+1)$. It is well-known [15.8, .10, .14] that this transformation makes the problem be nondegenerate.

The enlarged transportation problem now is

$$\text{minimize} \quad \sum_{i=1}^{m+1} \sum_{j=1}^{n+1} c_{ij}x_{ij} \tag{15.5}$$

$$\text{s.t.} \quad \sum_{j=1}^{n+1} x_{ij} = a_i' \quad \text{for} \quad i = 1,\ldots,m+1 \tag{15.6}$$

$$\sum_{i=1}^{m+1} x_{ij} = b_j' \quad \text{for} \quad j = 1,\ldots,n+1 \tag{15.7}$$

$$x_{ij} \geq 0 \quad \text{for} \quad i = 1,\ldots,m+1 \quad \text{and} \quad j = 1,\ldots,$$

$$n+1 \ . \tag{15.8}$$

For a nondegenerate problem a basic optimal solution has $m + n + 1$ positive x_{ij}'s ; all the rest are 0 . The cells in the basis form a tree. The dual to the problem in (15.5) - (15.8) is

$$\text{maximize} \sum_{i=1}^{m+1} u_i a_i + \sum_{j=1}^{n+1} v_j b_j \tag{15.9}$$

$$\text{s.t.} \ u_i + v_j \geq c_{ij} \quad \text{for} \quad i = 1,\ldots,m+1 \ ,$$

$$j = 1,\ldots,n+1 \tag{15.10}$$

where u_i and v_j are unconstrained variables. It is easy to show [15.14] that there is a one-parameter family of solutions to the dual problem. That is, if we fix any one of the u_i or v_j variables then all others are uniquely determined (for nondegenerative problems).

Because of the non-uniqueness of the dual variable solutions the standard parametric programming techniques of linear programming are not directly applicable. However, Srinivasan and the author showed in [15.25] that by defining the dual matrix (which is unique) with

$$d_{ij} = u_i + v_j \quad \text{for} \quad i = 1,\ldots,m+1 \ ,$$

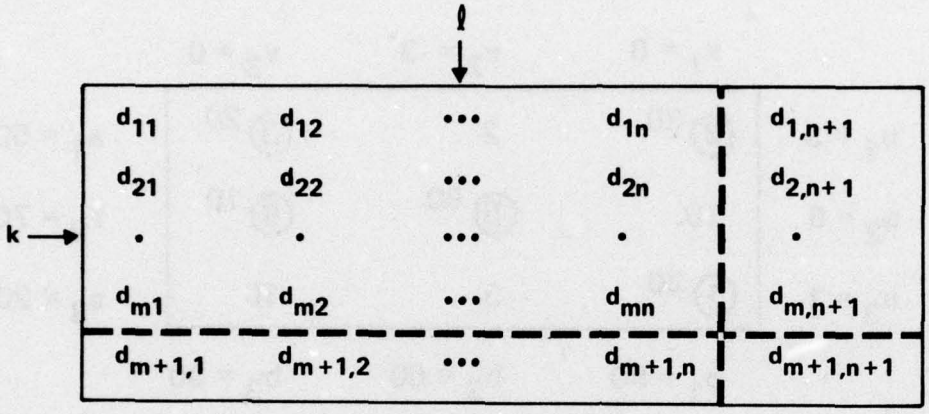
$$j = 1,\ldots,n+1 \tag{15.11}$$

a much richer and more detailed kind of parametric analysis was possible. What we showed was that, instead of having just the $m + n + 2$ dual variables u_i and and v_j of the problem in (15.9) and (15.10), it is possible to interpret each of the $(m+1)(n+1)$ numbers d_{ij} in (15.11) as "dual evaluators" entirely similar to ordinary dual variables. Some of these results are

summarized in Figure 15.1. (The complete results appear in [15.25] p. 250.) In Figure 15.1(a) the $(m+1)x(n+1)$ dual matrix is shown divided into four areas by dotted lines. Also a special row k which is a row in which a basis cell of the optimum solution occurs in column $n+1$, and a special column ℓ which is a column in which a basis cell of the optimum solution occurs in row $m+1$. If we then define rim operators which replace a_i by $a_i + \delta$ or $a_i - \delta$ and b_j by $b_j + \delta$ or $b_j - \delta$, and alter the primal and dual solutions accordingly to maintain optimality, we see that we can predict the effect on the optimum shipping cost by using the entries of Figure 15.1(b). Note that the prediction depends on whether Σa_i is $<$, $=$, or $>$ Σb_j. Also there is a definite range over which the prediction holds true, as in ordinary parametric analysis for linear programming. The details of ranging are given in [15.25].

A specific 3×3 numerical example is shown in Figure 15.2(a), together with its optimal primal and dual solution. The dual matrix for the problem appears in Figure 15.2(b) and various predictions as to changes in optimal shipping cost as the result of changes in rims appear in Figure 15.3. Note that predictions of the effects of changes in just an a_i or a b_j are given by d_{ij} entries in column $n+1$ or row $m+1$, while predictions of the effects of changes in both an a_i and a b_j are given by d_{ij} entries in the $m \times n$ matrix in the upper left-hand corner of Figure 15.2(b). In particular, note the fourth row of Figure 15.3 where by adding 1 to a_3 and 1 to b_2 we can ship one more ton while reducing the total shipping cost by 2. This is the so-called transportation paradox which occurs when rim changes permit "shipping more for less." Note that the dual matrix permits a complete prediction and explanation of the "paradox."

One of the most interesting numbers in the dual matrix is $d_{m+1,n+1}$ which can be shown to be the "downward marginal cost" or the "cost of the most expensive unit shipped." It can be shown to satisfy

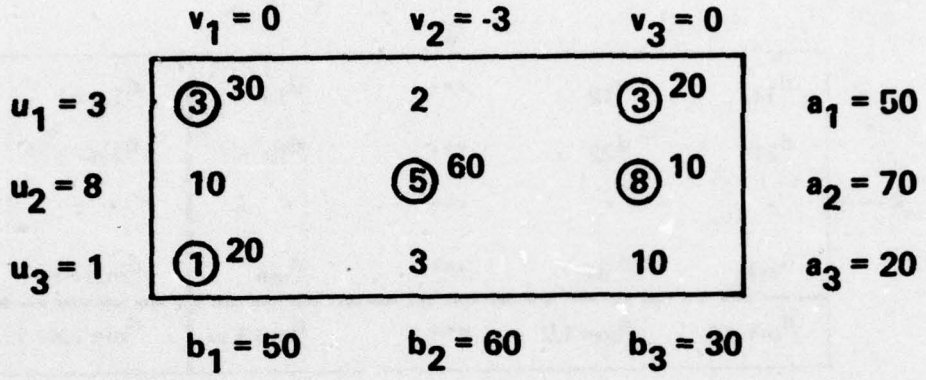


(a)

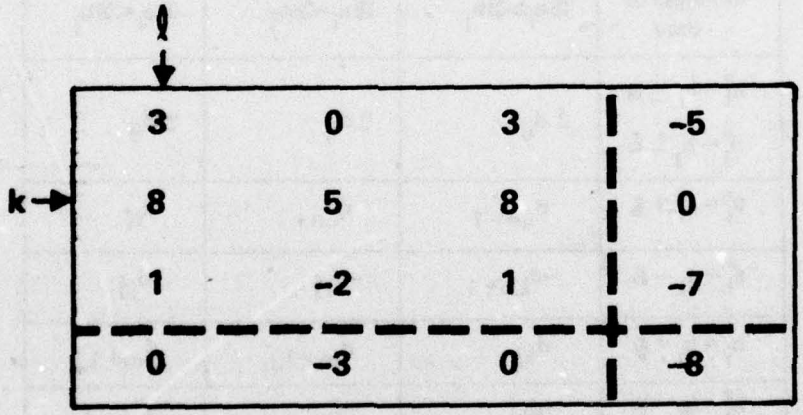
Changes in data	$\Sigma a_i > \Sigma b_j$	$\Sigma a_i = \Sigma b_j$	$\Sigma a_i < \Sigma b_j$
$a_i^* = a_i \pm \delta$ $b_j^* = b_j \pm \delta$	$\pm d_{ij}$	$\pm d_{ij}$	$\pm d_{ij}$
$a_i^* = a_i + \delta$	$d_{i,n+1}$	$d_{i,n+1}$	$d_{i\ell}$
$a_i^* = a_i - \delta$	$-d_{i,n+1}$	$-d_{i\ell}$	$-d_{i\ell}$
$b_j^* = b_j + \delta$	d_{kj}	$d_{m+1,j}$	$d_{m+1,j}$
$b_j^* = b_j - \delta$	$-d_{kj}$	$-d_{kj}$	$-d_{m+1,j}$

(b)

Figure 15.1 - Dual matrix and dual variable predictors for data changes.



(a)



(b)

Figure 15.2 - A numerical example

a_i changes	b_j changes	Total Cost Change Prediction
$a_1 = 51$	--	-5
--	$b_2 = 61$	-3
$a_2 = 71$	$b_1 = 51$	+8
$a_3 = 21$	$b_2 = 61$	-2
$a_2 = 69$	$b_1 = 49$	-8

Figure 15.3 - Predictions of changes in optimal shipping costs.

$d_{m+1,n+1} = -d_{k\ell}$ and its effect is illustrated in rows 3 and 5 of Figure 15.3.

Besides rim operators, Reference [15.25] discusses cost operators which simultaneously change costs c_{ij} and alter the primal and dual solutions to maintain optimality. Predictions for the effects of such changes are given by the corresponding x_{ij}'s. For both rim and cost operators, Reference [15.25] provides a complete discussion of the following.

(a) The maximum extent of the data change permitted while retaining the same optimal basis--these are basis preserving operators.

(b) The necessary basis changes to permit further data changes--these are global operators.

(c) The extension of both rim and cost operators to permit several rims or several costs to be changed simultaneously--these are area operators.

Space does not permit fuller discussion of these but an interested reader may consult [15.25].

Our purpose in developing the operator theory of parametric programming was not to carry out a mere technical exercise but to provide a tool for the solution of other problems. As a by-product we also found that operator theory gives new algorithms for the solution of the transportation problem! We discuss several of the applications of operator theory that have so far been made.

(A) Optimal Growth Paths. In [15.26] V. Srinivasan and the author studied the growth path of a transportation system, that is, the way "that the total cost, supplies and demands at origins and destinations, and the pattern of optimal shipments change when the total volume handled in the system increases subject to lower bounds on each origin's supply and each destination's demand." The kinds of questions that one might answer by solving this problem are: (1) How best to expand production facilities located at different places; (2) How to react to a growing market; (3) When to open or close factories or warehouses to react to a changing market. The technique employed in [15.26] for the solution of the problem was first to derive a modified problem having an extra row and column and then to apply a rim operator that simultaneously reduces the rim amounts on these new rows and columns. By applying this special rim operator using the algorithm described there we were able to trace out the optimal growth path as additional resources were invested in the problem. More details, including a numerical example, are given in [15.26].

(B) Assigning Uses to Sources. Consider a transportation problem with the added requirements that every market (use) must be supplied from a single warehouse (source). Such problems arise (1) in the scheduling of jobs to a group of identical computers where the added condition is that a job should be completely done on a single machine and (2) in the assignment of supervisory tasks to managers where the added condition is that a manager should either be assigned complete responsibility or no responsibility for any given super-

visory task. From a technical point of view, what this means is that we want the lowest cost feasible solution that has exactly one basis cell in each column except the last. In [15.27] V. Srinivasan and the author describe a branch and bound method for solving this problem that makes use of cost operators to drive cells into and out of the basis as required during the course of the branch and bound search. Thus, to drive a cell into the basis it is only necessary to apply a cost operator to drive it to a very small value (say - M) and to drive a cell out of the basis we apply another cost operator to make it very large (say M). Usually, only a few primal pivots are necessary to do either of these, so the work of applying the operator is much less than solving the transportation problem from scratch. Further details are given in [15.27]. The algorithm has been generalized by V. Balachandran [15.1]. His generalization is briefly discussed in Section 15.5.

(C) Solving Scheduling Problems. In [15.30] V. Srinivasan and the author made use of cost operators to aid in the branch and bound solution of two scheduling problems, the traveling salesman problem, and the decision critical path problem. In the case of the traveling salesman problem, we used a modification of the subtour elimination method in which we use the dual matrix to obtain at low computational cost weak lower bounds for decision nodes, and then used cost operators to drive cells into or out of the basis as search proceeded up and down the search tree. A similar approach was used in solving the decision critical path problem.

(D) Cash Management. In [15.23] V. Srinivasan discusses a very interesting application of transportation problems to solve the cash management problem of a firm that has cash payments and securities coming due in each of several future months, a possible line of credit from a bank, and certain needs for cash in each of these months. His work was a reformulation of that of Y. Orgler [15.22]. The problem is to determine what is the optimum schedule of liquidating the securities, perhaps before their due date, in order to meet the cash needs. Also, if there is surplus cash, how should it

be invested in securities to meet debts arising still further in the future? Again, the dual matrix and operator theory are useful in solving the problem. In fact, in his numerical example he used a negative rim operator to determine the optimal balance to keep in the bank.

(E) Choosing Modes of Transportation. There are several applications in which we have a transportation problem with more than one objective function. For instance, in many real transportation problems a shipper may want to minimize the time as well as the cost of shipping goods. These usually are contradictory objectives so that it is not possible to achieve both simultaneously. Instead, it is perhaps best to follow the suggestion of Geoffrion [15.15] and trace out a Pareto optimal surface in the time-cost space. In [15.31] V. Srinivasan and the author solved the problem of choosing different modes of transportation (rail, truck, airplanes) to try to minimize total costs and average shipment times. We used operator theory to trace out the Pareto optimal surface in time-cost space.

(F) Cost Operator Algorithm. The operators previously discussed permit the derivation of new algorithms for solving transportation and assignment problems. One of these is the cost operator algorithm which V. Srinivasan and the author described in detail in [15.32]. In this algorithm we first find the minimum entry in each row and subtract that amount from each row entry; then we find the minimum entry in each column and subtract that amount from each column entry. This gives an equivalent problem with at least one 0 cost in each row and column. Next we find a primal feasible starting solution by any convenient method. If all the basis cells of this starting solution have 0 cost, then it is optimal. If not, let A be the subset of basis cells having positive cost. We change the costs c_{ij} for $(i,j) \in A$ to $c^*_{ij} = 0$ and now the solution is optimal for the altered problem. We now apply cost operators to drive the costs for $(i,j) \in A$ back to their original values. In [15.32] we prove that this algorithm converges, even for a primal degenerate

problem. We also compare computational times with this code with our 1971 primal code and find that the cost operator algorithm takes about 3-4 times longer than the primal method for transportation problems and 1.5 - 2 times longer for assignment problems. In [15.32] we also derive theoretical bounds on the number of pivot steps needed and show that these bounds are considerably stronger than the corresponding bounds derived by Ford and Fulkerson [15.11] for their dual maximum flow method.

15.3 Operator Theory of Parametric Programming for the Generalized Transportation Problem

In a series of papers [15.4, .5, .6, .7] V. Balachandran and the author have extended the operator theory of parametric programming to the generalized transportation problem. We found that most of the same results hold true in the generalized case, and in addition some new results can be obtained. We briefly survey some of these results and then discuss applications of the model.

In order to define the model we use the familiar machine loading problem for its interpretation. Suppose we have m machine types $i = 1,\ldots,m$ and n product types $j = 1,\ldots,n$. We define the following.

x_{ij} = the amount of product type j to be produced on machine type i

c_{ij} = the unit cost of such production

e_{ij} = the production time required by machine i to produce one unit of product j

a_i = the total time available for machine i

b_j = the number of units of j needed

We want to make the required production at minimum total cost. Hence, we have the following generalized transportation problem.

$$\text{minimize} \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} e_{ij} x_{ij} \leq a_i \qquad (15.12)$$

$$\sum_{i=1}^{m} x_{ij} = b_j$$

$$x_{ij} \geq 0$$

Note that the only difference from the ordinary transportation problem (15.1) - (15.3) is constraint (15.12) which has the e_{ij} coefficients and a $\leq$ sign. However, this change makes the problem considerably more difficult to solve.

As in the previous case, we add another row and column by defining the following

$$c_{i,n+1} = 0 \text{ , and } e_{i,n+1} = 1 \text{ for } i = 1,\ldots,m+1 \qquad (15.13)$$

$$c_{m+1,j} = M \text{ , and } e_{m+1,j} = 1 \text{ for } j = 1,\ldots,n$$

where M is a large number. We also define

$$a_{m+1} = M$$

Thus, the last column acts as a slack column and there is no constraint on the $x_{i,n+1}$ variables. The last row permits an easy starting solution to be found and because of (15.13) optimality requirements will drive $x_{m+1,j} = 0$ for $j = 1,\ldots,n$, if possible.

The new problem is defined by

$$\text{maximize} \quad \sum_{i=1}^{m+1} \sum_{j+1}^{n+1} c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n+1} e_{ij}x_{ij} = a_i$$

$$\sum_{i=1}^{m+1} x_{ij} = b_j$$

$$x_{ij} \geq 0$$

For a nondegenerate problem a basic optimal solution will have $m + n + 2$ positive x_{ij}'s and the rest are 0 . However, the topology of the solution is more complicated than in the previous case. In the present case it is a forest (union) of 1-trees where we define a 1-tree to be a tree with one additional edge.

The dual problem is

$$\text{maximize} \quad \sum_{i=1}^{m+1} u_i b_i + \sum_{j=1}^{n} v_j b_j \tag{15.14}$$

$$\text{s.t.} \quad u_i e_{ij} + v_j \leq c_{ij} \quad \text{for} \quad i = 1,\ldots,m+1$$

$$\text{and} \quad j = 1,\ldots,n \tag{15.15}$$

In this case the solution for the dual variables is unique. In addition, it can easily be proved that $u_i \leq 0$ at the optimum. If $c_{ij} \geq 0$ and $e_{ij} \geq 0$ it can also be proved that $v_j \geq 0$ at the optimum.

The <u>dual matrix</u> for the optimum solution is defined to have

$$d_{ij} = e_{ij}u_i + v_j \quad \text{for} \quad i = 1,\ldots,m+1 \quad \text{and} \quad j = 1,\ldots,n$$

As in the previous case if we define rim operators that replace a_p by $a_p + e_{pq}\delta$ and b_q by $b_q + \delta$ we can show (see [15.4, .7]) that d_{pq} predicts, over a certain range, the change in total cost of the optimal

solution to the changed problem. Note again that instead of the $m + n + 1$ dual variable evaluators that one would normally expect from the dual problem (15.14) - (15.15) we find $(m+1)n$ such evaluators, each with a range and rules for proceeding from optimum basis and primal and dual solution to the next as the operator is applied for larger and larger $|\delta|$. There is also the possibility of a "production paradox" similar to the "transportation paradox" of the ordinary problem. In [15.5, .7] we describe how cost operators can be applied to the problem with results similar to those in the ordinary transportation problem case. Finally, in [15.6, .7] we describe a new type of operator, the weight operator, that can be applied to a generalized transportation problem with its optimal solution and drive it to another such problem that differs only in its weights from the previous problem. All of these operators permit the solution of problems that do not directly fit the generalized transportation framework. We proceed to discuss some of the applications so far made.

(A) Uses to Sources for the Generalized Problem. In [15.1] V. Balachandran generalized the results of [15.27] to apply to the generalized transportation problem. He used as a specific example the problem of optimal job assignment in a computer network. In this case the e_{ij} coefficients measure how much time the ith computer needs to work on job j and it is possible in this model to have the efficiency of various computers in the network to be different for a given job.

(B) Stochastic Generalized Transportation Problem. In [15.2] V. Balachandran utilized operator theory for the solution of a generalized transportation problem in which the demands, the b_j's, vary stochastically. He first derived the Kuhn-Tucker conditions for the problem and then showed with successive applications of area rim operators how these conditions could be satisfied. His work is a generalization of that of Garstka [15.13] for the ordinary transportation problem who suggested resolving the problem at each iteration.

(C) A Cash Management Example. In [15.23] V. Srinivasan mentioned that a more realistic version of the cash management example could be obtained by using a generalized transportation model. Specifically, if a bill due in period 1 is paid with cash from period 1, it has an e factor of 1.02 because of a 2% discount for cash. However, if the bill from period 1 is paid (late) with cash from period 2, its e factor is 0.98 because of a penalty for late payment. Figure 15.4 shows a simple two-period example with sources of cash being cash coming in and securities coming due, and needs being cash payments coming due and accounts payable. It is assumed there is a 2% discount for cash, and that cash can be invested for 1 period at a 1% return. The securities have various returns indicated by the e_{ij} factors in the upper left-hand corner of each cell. The costs are the negatives of the profit realized from the transaction in the cell. The positive x_{ij}'s are shown in the upper right-hand corner of basis cells and the objective function is the negative of the profit. Since we minimize negative profits, this is the same as maximizing profits. The objective value is -3.916 so that the rate of return is 3.916/87 = 0.045 on cash and securities per period. The optimum dual matrix is shown in Figure 15.5. From the last column we see that more cash in period 2 is of no value but is worth 0.0097 per $ in period 1. Also, additional securities in period 1 are worth 0.146 per $ and 0.138 per $ in period 2. Note also that if we can increase cash inflow in period 1 from 30 to 31.01 and cash payments in period 2 from 22 to 23 we can increase profits by 0.0098. On the other hand, if we can increase cash income in period 2 from 35 to 36 and increase cash payments in period 1 from 15 to 16 we will decrease profits by 0.0097. Many other interpretations of dual evaluators can be deduced from this dual matrix.

15.4 The Time (Bottleneck) Transportation Problem

The time transportation problem, sometimes also called the bottleneck problem, differs from an ordinary transportation problem only in its objective function. Using the same notation as in Section 15.2 it can be

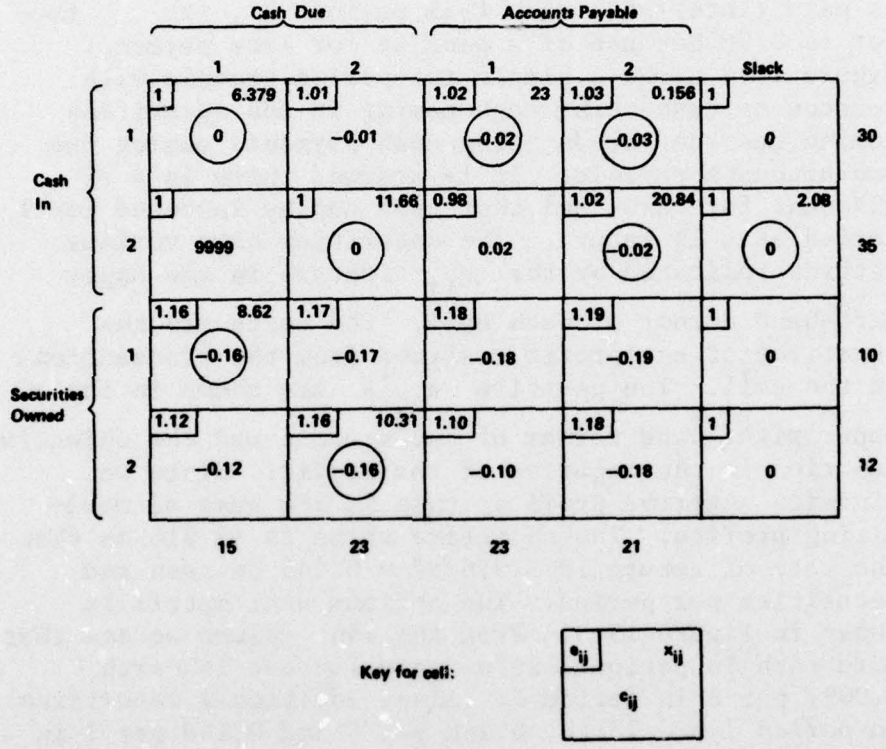


Figure 15.4 - A cash management example where the optimum value is −3.916.

0	-0.0098	-0.02	-0.03	-0.0097
0.0097	0	-0.01	-0.02	0
-0.16	-0.171	-0.182	-0.194	-0.146
-0.145	-0.16	-0.168	-0.183	-0.138
-0.0097	0	-0.01	-0.02	0

Figure 15.5 - Optimum dual matrix for cash management example.

stated as

$$\text{minimize} \quad \max_{x_{ij}>0} c_{ij} \tag{15.16}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = a_i$$

$$\sum_{i=1}^{m} x_{ij} = b_j$$

$$x_{ij} \geq 0$$

where we also assume $\Sigma a_i = \Sigma b_j$. Several solution procedures have been proposed by Szwarc [15.34], Hammer [15.17, .18], and Garfinkel and Rao [15.12]. V. Srinivasan and the author are preparing a comparison of these methods [15.33]. Specifically, we find a close affinity to operator theory of the methods of Swarc and Hammer. We have programmed these Szwarc methods as well as the cost operator method of our own

for the problem. Some computational experience with Szwarc's method is given in the next section.

Although the time transportation problem has only rarely been applied in practice, it seems to the author that it is a very attractive model for applications, especially in cases where adding together costs from different cells is not necessarily meaningful as in Example (B) below. In any case the existence of fast computer codes may encourage its application in the future. We conclude with two examples of such applications.

(A) Shipping Perishable Goods to Market. One of the earliest suggested applications of the time transportation problem is that of shipping perishable goods from warehouses to markets, where c_{ij} is the time to send the goods from W_i to M_j. Then the quantity

$$\underset{x_{ij}>0}{\text{maximum}} \; c_{ij}$$

measures the longest in-transit time of any shipment. Clearly, the objective function (15.16) means that we want to find a shipping pattern whose longest transit time is minimal.

(B) Personnel Assignment Problem. In [15.9] Charnes, Cooper, Niehaus and Stedry proposed several different problem formulations to take into account the multiple objectives that organizations may consider in assigning men to jobs. In [15.28] V. Srinivasan and the author showed that their reformulation for static multi-attribute models which used the Chebyshev metric could be reformulated as a bottleneck assignment problem. The bottleneck assignment problem is a square time transportation problem in which all $a_i = b_j = 1$. We will illustrate this in terms of a simple example.

We use the notation of [15.9, .28] by defining

x_{is} = amount of job i assigned to individual s

r_{ij} = amount of jth attribute required by job i

a_{sj} = amount of jth attribute possessed by individual s

Then if we define

$$c_{is} = \max_{j} \left| r_{ij} - a_{sj} \right| \text{ for all } i \text{ and } s \quad (15.17)$$

we can formulate the bottleneck personnel problem as

$$\text{minimize} \quad \max_{x_{is}>0} c_{is} \quad (15.18)$$

$$\text{s.t.} \quad \sum_{i} x_{is} = 1 \quad \text{for all} \quad s \quad (15.19)$$

$$\sum_{s} x_{is} = 1 \quad \text{for all} \quad i \quad (15.20)$$

$$x_{is} \geq 0 \quad \text{for all} \quad i \quad \text{and} \quad s \quad (15.21)$$

Then, because the bottleneck transportation or assignment problem always has integer solutions if the original rim data are in integers, we also know that the solution will satisfy $x_{is} = 0$ or 1 as well.

Note that the definition in (15.17) of the objective function coefficients indicates that c_{is} measures the maximum deviation above or below the goal of any attribute a when a given individual s is to be assigned to job i. Then the solution of (15.18) - (15.21) yields an assignment of men to jobs so that the overall maximum deviation from the goal of any assignment of a man to a job is minimized. This means that every assignment of a man to a job is no worse than the optimum bottleneck value.

A specific example of such an assignment problem having 5 jobs and 5 men is shown in Figure 15.6. The solution to the problem considered as a bottleneck problem is shown in Figure 15.7(a). For comparison, its solution as an ordinary assignment problem is shown in

Jobs \ Men	1	2	3	4	5	
1	7	1000	12	6	0	1
2	1000	5	20	1000	1000	1
3	7	1000	1000	11	90	1
4	6	12	13	1000	6	1
5	12	25	0	1000	1000	1
	1	1	1	1	1	

Figure 15.6 - An assignment problem.

I	J	C(I,J)
1	4	6
2	2	5
3	1	7
4	5	6
5	3	0
		24

(a) Bottleneck Solution
(Bottleneck cost = 7)

I	J	C(I,J)
1	5	0
2	2	5
3	4	11
4	1	6
5	3	0
		22

(b) Assignment Solution
(Largest cost = 11)

Figure 15.7 - Two solutions to assignment problem.

Figure 15.7(b). Note that the solution of Figure 15.7(b) has two men perfectly matched while that of Figure 15.7(a) has only one match. However, the worst match in Figure 15.7(a) has a penalty cost of 7 compared with the worst penalty of 11 in Figure 15.7(b). Here is an application for which the bottleneck metric seems much better suited than the ordinary one.

15.5 Computational Comparisons of Algorithmic Efficiency

Since the mid-Fifties when computers became available, computational tests of various algorithms have been made and reported in the literature. Unfortunately, most of these reports were given in the form "algorithm X runs K times faster than algorithm Y." Thus, computational speeds became a matter of forklore to be passed by word of mouth. In recent years the standards for reporting results have improved but comparisons are still made on a given coding of a given algorithm by one group with a previous coding of a different algorithm by another group. For instance, here are some of the published reports of computational comparisons between the primal (MODI) method and primal-dual (Ford-Fulkerson) method.

1. In 1963 Ford and Fulkerson [15.11] reported that the stepping-stone method [15.8] takes about twice as long as their method to solve a transportation problem.

2. In 1972, Hatch, Nauta, and Pierce [15.19] reported that their DSAI-MAXFLOW algorithm was from 2-24 times faster than the Ford-Fulkerson maximum flow algorithm.

3. In 1973, Srinivasan and the author reported that our 1971 primal code solved either transportation or assignment problems in about the same time as the Hungarian assignment problem algorithm requires for the same size assignment problems only; see [15.29].

4. In 1974, Glover, Karney, Klingman, and Napier reported that their primal code was 2-20 times

faster than Fulkerson's out-of-kilter dual method [15.16].

5. In 1974 Srinivasan and the author reported [15.33] that Szwarc's algorithm for the time transportation problem runs approximately twice as fast as their 1971 primal ordinary transportation code in solving problems of the same size.

With the exception of 3 and 5 above these comparisons were made between codes written by different people and run on different machines. However, it is easy to show that running times depend upon, at least, the following factors.

(a) The mathematical algorithm used and its precise description

(b) The computer science methods used, such as list structures, search techniques, and so on

(c) The way test problems are generated

(d) The computer language used

(e) The "cleverness" of the programmer

(f) The compiler used--some compilers execute much faster than others, even on the same machine

(g) Computing machine used

From this we can conclude that timing tests between algorithms will probably reach different conclusions when performed by different groups at different locations at different times. It is also clear from the timing tests reported above and those soon to be released, that the sophistication and size of problems that can be solved are both increasing rapidly.

The author can also report his personal experience that friendly competition and cooperation among groups attacking the same kinds of problems is highly beneficial to everyone.

There are still a number of questions to be answered, among which are the following.

1. For what kinds of problems is the primal (MODI) method necessarily better (or worse) than the primal-dual method?

2. Are there other algorithms (such as [15.32]) that are superior to either of these?

3. For any given algorithm, what is the best way of organizing the actual code, that is, what are the best list structures, the best search techniques, best labelling methods, and so on?

4. Are there other experimental results, such as the minimum cost effect of [15.28], that help to characterize problem difficulty?

15.6 Conclusions

Research on the mathematical presentation and algorithmic implementation of transportation codes has been and will continue to be intensive. Relative judgments as to algorithmic efficiencies change as new mathematical or implementation ideas are developed.

One clear conclusion is that the old paradigm that a mathematician develops an algorithm and a programmer writes the code must be abandoned. Instead, work on the problems discussed in this paper clearly shows that the best codes are developed by intertwining mathematical and algorithm ideas together with experimental results. The programmers must also be mathematicians and vice versa.

George B. Dantzig has recently recommended the development of "centers of computational excellence" at various places in the country, with each center taking responsibility for certain algorithm areas. Such a center should provide the following.

(a) Test problems both of the randomly generated and "real" varieties

(b) Facilities for testing codes written by other people

(c) Codes for users

These centers would require substantial financial support since algorithm testing requires many hours of fast computer time. However, the results reported on so far already show the potentially large benefits available from such in-depth studies of computer algorithms. The author heartily endorses and supports the idea of the establishment of such centers.

References

[15.1] Balachandran, V. (1972). An integer generalized transportation problem for optimal job assignment in computer networks. Management Science Research Report No. 308, Carnegie-Mellon University, (November).

[15.2] Balachandran, V. (1973). The stochastic generalized transportation problem--an operator theoretic approach. Management Science Research Report No. 311, Carnegie-Mellon University, (February).

[15.3] Balachandran, V., and G. L. Thompson (1973). The four index algorithm for the generalized transportation problem. Management Science Research Report No. 310, Carnegie-Mellon University, (March).

[15.4] Balachandran, V., and G. L. Thompson (1975a). An operator theory of parametric programming for the generalized transportation problem. I. Basic theory. Naval Res. Logist. Quart. 22 79-100.

[15.5] Balachandran, V., and G. L. Thompson (1975b). An operator theory of parametric programming for the generalized transportation problem. II. Rim, cost and bound operators. Naval Res. Logist. Quart. 22 101-125.

[15.6] Balachandran, V., and G. L. Thompson (1975c). An operator theory of parametric programming for the generalized transportation problem. III. Weight operators. Naval Res. Logist. Quart. 22 297-315.

[15.7] Balachandran, V., and G. L. Thompson (1975d).

An operator theory of parametric programming for the generalized transportation problem. IV. Global operators. Naval Res. Logist. Quart. 22 317-339.

[15.8] Charnes, A., and W. W. Cooper (1961). Management Models and Industrial Applications of Linear Programming. I and II. Wiley.

[15.9] Charnes, A., W. W. Cooper, R. J. Niehaus, and A. Stedry (1969). Static and dynamic assignment models with multiple objectives and some remarks on organizational design. Management Sci. 15 365-375.

[15.10] Dantzig, G. B. (1963). Linear Programming and Extensions. Princeton University Press.

[15.11] Ford, L. R., Jr., and D. R. Fulkerson (1962). Flows in Networks. Princeton University Press.

[15.12] Garfinkel, R. S., and M. R. Rao (1971). The bottleneck transportation problem. Naval Res. Logist. Quart. 18 465-472.

[15.13] Garstka, S. J. (1970). Computation in stochastic programs with recourse. Management Science Research Report No. 204, Carnegie-Mellon University, (April).

[15.14] Gaver, D. P., and G. L. Thompson (1973). Programming and Probability Models in Operations Research. Brooks/Cole.

[15.15] Geoffrion, A. M. (1967). Solving bicriterion mathematical programs. Operations Res. 15 39-54.

[15.16] Glover, F., D. Karney, D. Klingman, and A. Napier (1974). A computational study on start procedures, basis change criteria, and solution algorithms for transportation problems. Management Sci. 20 793-813.

[15.17] Hammer, P. L. (1969). Time-minimizing transportation problems. Naval Res. Logist. Quart. 16

345-357.

[15.18] Hammer, P. L. (1971). Communication on "the bottleneck transportation problem" and some remarks on "the time transportation problem." Naval Res. Logist. Quart. 18 487-490.

[15.19] Hatch, R. S., F. Nauta, and M. B. Pierce (1972). Development of generalized network flow algorithms for solving the personnel assignment problem. Decision Systems Associates, Inc. (April).

[15.20] Klingman, D., A. Napier, and J. Stutz (1974). Netgen: a program for generating large scale capacitated assignment, transportation, and minimum cost flow network problems. Management Sci. 20 814-821.

[15.21] Kuhn, H. W. (1956). The Hungarian method for the assignment problem. Naval Res. Logist. Quart. 3 253-258.

[15.22] Orgler, Y. (1970). Cash Management: Methods and Models. Wadsworth.

[15.23] Srinivasan, V. (1974). A transshipment model for cash management decisions. Management Sci. 20 1350-1363.

[15.24] Srinivasan, V., and G. L. Thompson (1972a). Accelerated algorithms for labelling and relabelling of trees, with applications to distribution problems. J. Assoc. Comput. Mach. 19 712-726.

[15.25] Srinivasan, V., and G. L. Thompson (1972b). An operator theory of parametric programming for the transportation problem, Parts I and II. Naval Res. Logist. Quart. 19 205-252.

[15.26] Srinivasan, V., and G. L. Thompson (1972c). Determining optimal growth paths in logistics operations. Naval Res. Logist. Quart. 19 575-599.

[15.27] Srinivasan, V., and G. L. Thompson (1973a). An algorithm for assigning uses to sources in a special

class of transportation problems. Operations Res. 21 284-295.

[15.28] Srinivasan, V., and G. L. Thompson (1973b). Alternate formulations for static multi-attribute assignment models. Management Sci. 20 154-158.

[15.29] Srinivasan, V., and G. L. Thompson (1973c). Benefit-cost analysis of coding techniques for the primal transportation algorithm. J. Assoc. Comput. Mach. 20 194-213.

[15.30] Srinivasan, V., and G. L. Thompson (1973d). Solving scheduling problems by applying cost operators to assignment models. in S. E. Elmaghraby (ed.) Symposium on the Theory of Scheduling and its Applications. Lecture Notes in Economics and Mathematical Systems 86 Springer. 399-425.

[15.31] Srinivasan, V., and G. L. Thompson (1974a). Determining cost vs. time Pareto-optimal frontiers in multi-modal transportation problems. Management Science Research Report No. 292, Carnegie-Mellon University, (Revised January).

[15.32] Srinivasan, V., and G. L. Thompson (1974b). Cost operator algorithms for the transportation problem. Management Science Research Report No. 344, Carnegie-Mellon University, (August).

[15.33] Srinivasan, V., and G. L. Thompson (1975). Algorithms for minimizing total cost, bottleneck time and bottleneck shipment in transportation problems. Management Science Research Report No. 353, Carnegie-Mellon University, (January).

[15.34] Szwarc, W. (1971). Some remarks on the time transportation problem. Naval Res. Logist. Quart. 18 473-485.

Chapter 16

A SURVEY OF APPLICATIONS OF INTEGER AND COMBINATORIAL PROGRAMMING IN LOGISTICS*

Jeremy F. Shapiro
Massachusetts Institute of Technology

16.1 Introduction

There are a number of definitional ground rules to be established before we enter into our survey. First, an application is taken to be a study in which concern over a real-world problem caused the formulation of an integer or combinatorial programming model, the collection of data for this model, and the calculation of numerical solutions using a computer. This is in contrast to studies in other social science fields where mathematical models are used to obtain qualitative insights without necessarily requiring data and numerical calculations.

A second ground rule is to agree that we will not try to define logistics, but rather to consider specific illustrative applications that most of us would agree address logistics problems. These applications are chosen from the functional areas of distribution, location, scheduling, production/inventory control, communications and reliability.

Another reason for considering illustrative applications is that the number of applications is enormous and a comprehensive survey is not possible. Our purpose instead is to discuss by example the underlying principles used in these applications. The principles are derived from the synergism that exists between mathematical programming theory as it relates to algorithms, the construction and use of computer systems, and the institutional aspects of the applications themselves.

---

*The preparation of this chapter was supported in part by the Army Research Office-Durham under Contract DA HC04-70-C-0058 with the Massachusetts Institute of Technology.

16.2 Discrete Programming

In mathematical terms, the most general statement of the class of mathematical programming models we will discuss is the following. The object is to minimize the quantity $f(x)$ where the vector x is chosen from a finite or denumerable set X contained in a finite dimensional space, say R^n. The set X may be given implicitly or defined explicitly by a set of constraint functions including integrality restrictions on the variable values. Discrete programming differs from nonlinear programming in that differential methods cannot be used directly to analyze the objective and constraint functions. Moreover, convex combinations of solutions from X may not themselves be points in X and therefore linear programming approximations may be inexact.

Within the class of discrete programming problems there are two overlapping subclasses: integer programming and combinatorial programming problems. We can think of integer programming problems as being of the form

$$\text{minimize} \quad f_1(x) + f_2(y)$$

$$\text{s.t.} \quad A_1(x) + A_2(y) \geq b$$

$$x \geq 0 \ , \ y \geq 0 \ \text{and integer}$$

where usually $A_2(y) = A_2 y$, that is, the function $A_2(y)$ is a linear function, and slightly less often $f_2(y) = f_2 y$. For a system problem such as this one, one uses integer programming system theory including number theory and branch and bound (for example, Geoffrion and Marsten (1972), Gorry, Northup and Shapiro (1973)).

By contrast, combinatorial programming problems have a less explicit mathematical statement. They contain network optimization problems as substructures including shortest route, maximal flow, minimum spanning tree and minimum cost flow problems. All of these network optimization problems can be solved by "good" algorithms

which means algorithms with a number of steps upper bounded by a polynomial in the parameters of the problem (Edmonds (1971), Karp (1972)). An algorithm is not "good" if it is possible for the algorithm to require on some problems a number of steps that grows exponentially with the parameters of the problem. "Good" algorithms are good in a practical as well as theoretical sense and network optimization problems of significant size can often be solved in a matter of a few seconds on large-scale computers (Glover et al. (1974)).

There are other relatively simple combinatorial optimization problems which appear as subproblems in applications. These include simple covering and matching problems (Garfinkel and Nemhauser (1972)), discrete deterministic dynamic programming problems (Wagner (1969)), and others. Although "good" algorithms may not exist for these problems, they are often easy to solve relative to the complex combinatorial programming problems found in practice.

Specifically, the combinatorial programming models arising in logistics applications are often a synthesis of several similar or different problems of the above types, plus complicating constraints or relations. Practically all of these problems can be formulated as integer programming problems, but often the special structure of the problem is lost. A good example of this is the symmetric traveling salesman problem for which there is an integer programming formulation with approximately 2^n constraints, where n is the number of cities to be visited (Held and Karp (1970)). The majority of these constraints, however, describe a minimal spanning tree problem, and Held and Karp (1970, 1971) exploit this structure in a special purpose algorithm for the traveling salesman that involves the solution of an effective n constraint approximation of the problem.

The choice of an integer programming or combinatorial programming formulation of a discrete optimization problem is closely related to the choice one must make between a general purpose or special purpose algorithm for the given problem. Unfortunately, this choice cannot always be made as definitively as it can be for

the traveling salesman problem. The conflict can be resolved in large part, however, by the modular design of integer programming and network optimization computer codes so that the synthesis required for a specific application can be made without a complete setup. As we shall see, the synthesis of a model from its component parts can be effected by the application of dual or price directive decomposition methods of mathematical programming. Decomposition can also be effected by resource directive methods, but this approach has found little if any application. See Lasdon (1970) for a discussion of these approaches.

Illustrative Application One. Multi-item Production Scheduling and Inventory Control (Lasdon and Terjung (1971)).

Consider a manufacturing system consisting of I items for which production is to be scheduled over T time periods. The demand for item i in period t is the nonnegative integer r_{it}; this demand must be met by stock from inventory or by production during the period. Let the variable x_{it} denote the production of item i in period t. The inventory of item i at the end of period t is

$$y_{it} = y_{i,t-1} + x_{it} - r_{it} \qquad t = 1,\ldots,T$$

where we assume $y_{i,0} = 0$, or equivalently, initial inventory has been netted out of the r_{it}. Associated with x_{it} is a direct unit cost of production c_{it}. Similarly, associated with y_{it} is a direct unit cost of holding inventory h_{it}. The problem is complicated by the fact that positive production of item i in period t uses up a quantity $a_i + b_i x_{it}$ of a scarce resource q_t to be shared among the I items. The parameters a_i and b_i are assumed to be nonnegative.

Lasdon and Terjung (1971) applied this model to the scheduling of automobile tires production. The scarce

resource in each period was machine capacity. The number of different items (tires) was approximately 400, and the planning horizon was approximately 6 periods.

This problem can be written as the mixed integer programming problem

Problem 16.1

$$\text{Find } v = \text{minimum} \sum_{i=1}^{I} \sum_{t=1}^{T} (c_{it}x_{it} + h_{it}y_{it}) \qquad (16.1a)$$

$$\text{s.t.} \sum_{i=1}^{I} (a_i\delta_{it} + b_i x_{it}) \leq q_t \qquad (16.1b)$$

$$t = 1,\ldots,T; \text{ for } i = 1,\ldots,I$$

$$\sum_{t=1}^{s} x_{it} - y_{is} = \sum_{t=1}^{s} r_{it} \qquad (16.1c)$$

$$s = 1,\ldots,T$$

$$x_{it} \leq M_{it}\delta_{it}, \quad t = 1,\ldots,T \qquad (16.1d)$$

$$x_{it} \geq 0, \quad y_{it} \geq 0 \qquad (16.1e)$$

$$\delta_{it} = 0 \text{ or } 1, \quad t = 1,\ldots,T$$

where $M_{it} = \sum_{s=t}^{T} r_{is}$ is an upper bound on the amount we would want to produce in period t. The constraints (16.1b) state that shared resource usage cannot exceed q_t. For simplicity, we have assumed a single resource to be shared in each production period. The model can clearly be used when there are K shared resources in each period. The constraints (16.1c) relate accumulated production and demand through period t to ending inventory in period t, and the nonnegativity of the y_{it} implies demand must be met and not delayed

(backlogged). The constraints (16.1d) ensure that $\delta_{it} = 1$ and therefore the fixed charge resource usage a_i is incurred if production x_{it} is positive in period t. Problem 16.1 is a mixed integer programming problem with IT zero-one variables, $2IT$ continuous variables and $T + 2IT$ constraints. For the application of Lasdon and Terjung, these figures are 240 zero-one variables, 480 continuous variables, and 486 constraints which is a mixed integer programming problem of significant size.

For future reference, define the set

$$N_i = \{(\delta_{it}, x_{it}, y_{it}) \ , \ t = 1,\ldots,T \mid \delta_{it}, x_{it}, y_{it} \text{ satisfy (16.1c), (16.1d), (16.1e)}\} \tag{16.2}$$

This set describes a feasible production schedule for item i ignoring the joint constraints (16.1b).

The integer programming formulation (16.1) is not effective because it fails to exploit the special structure of the sets N_i. This can be accomplished by dual (price directive) decomposition with proceeds as follows. Assign prices $u_t \geq 0$ to the scarce resources q_t and place the constraints (16.1b) in the objective function to form the Lagrangian.

$$L(u) = -\sum_{t=1}^{T} u_t q_t + \underset{(\delta_{it}, x_{it}, y_{it}) \in N_i}{\text{minimum}} \sum_{i=1}^{I} \sum_{t=1}^{T} \{(c_{it} + u_t b_i) x_{it} + u_t a_i \delta_{it} + h_{it} y_{it}\}$$

Letting

$$L_i(u) = \underset{(\delta_{it},x_{it},y_{it})\in N_i}{\text{minimum}} \sum_{t=1}^{T} \{(c_{it}+u_t b_i)x_{it} + u_t a_i \delta_{it} + h_{it} y_{it}\} \tag{16.3}$$

the Lagrangian function clearly separates to become

$$L(u) = -\sum_{t=1}^{T} u_t q_t + \sum_{i=1}^{I} L_i(u)$$

Each of the problems (16.3) is a simple dynamic programming shortest-route calculation for scheduling item i where the dual prices on shared resources adjust the costs as shown.

It is easily shown that $L(u)$ is a lower bound on the minimal objective function cost v in Problem (16.1). The best choice of prices is a vector u^* which provides the greatest lower bound; namely, a vector u^* that is optimal in the dual problem

$$w = \text{maximum } L(u) \quad \text{s.t. } u \geq 0 \tag{16.4}$$

where clearly $w \geq v$. The reason for this selection of prices is that if the maximal dual objective function value w equals the minimal primal objective function value v, then it is possible to solve (16.1) by calculation of $L_i(u^*)$ for each item i. Approximate equality between v and w obtains when the number of items I is significantly greater than the number of joint constraints (16.1b) in the planning problem.

The dual problem (16.4) can be solved in a number of ways. One algorithm is generalized linear programming, otherwise known as Dantzig-Wolfe decomposition (Lasdon (1970)). This is the approach taken by Lasdon and Terjung who, in addition, used the generalized upper bounding technique (Lasdon (1970)) to solve the linear programming subproblems that arise in the use of this

algorithm. Further discussion about generalized linear programming and duality is contained in Magnanti, Shapiro and Wagner (1973).

If there is a substantial duality gap between the primal problem (16.1) and the dual problem (16.4) (that is, if $v - w$ is a large positive number), then problem (16.1) becomes more difficult to solve. In this case, the dual decomposition approach needs to be combined with branch and bound (see Fisher, Northup and Shapiro (1974)). To the best of my knowledge, the model (16.1) has never been used to analyze a real-life logistics problem where the number of joint constraints (16.1b) is of the same order of magnitude as the number of items for which production is being scheduled and a large duality gap is likely.

Another application of combinatorial methods to production is contained in Müller-Merbach (1973). He considers a production system consisting of a hierarchy of assemblies to be merged into final products. The assembly process is described as a network for the purposes of analyzing explosion of material requirements and costs.

Illustrative Application Two. Warehouse Location and Multi-Commodity Distribution (Geoffrion and Graves (1973)).

In the previous application, we considered a discrete optimization problem for which the mixed integer programming formulation was inefficient because it failed to exploit special structure. We consider now an application in which mixed integer programming was successfully applied. The model used in the application is an example of a large class called location-allocation problems (see Lea (1973) for an extensive bibliography).

The application of Geoffrion and Graves involved a two-level distribution system with plants, each producing a number of different commodities to be shipped to warehouses from which wholesale customers are supplied. The decisions to be made were: (1) what warehouse sites should be used; (2) what should be the size of each warehouse; (3) what customers should be served by each warehouse; and (4) what is the optimal pattern of multi-commodity transportation flows?

Let i be the index for commodities, j the index for plants, k the index for possible warehouse sites and ℓ the index for customers. Define the variables $x_{ijk\ell}$ as the nonnegative amount of commodity i produced in plant j for delivery to customer ℓ via a warehouse at site k. Let the zero-one variable j_k determine whether $(z_k=1)$ or not $(z_k=0)$ a warehouse is constructed at location k. Let the zero-one variable $Y_{k\ell}$ determine whether $(Y_{k\ell}=1)$ or not $(Y_{k\ell}=0)$ customer ℓ is supplied from warehouse k.

The warehouse location and multi-commodity distribution problem can be written as the mixed integer programming problem

Problem 16.5

$$\text{minimize} \quad \sum_{ijk\ell} c_{ijk\ell}x_{ijk\ell} + \sum_{k} \{f_k z_k + v_k \sum_{i\ell} d_{i\ell} y_{k\ell}\} \tag{16.5a}$$

$$\text{s.t.} \quad \sum_{k\ell} x_{ijk\ell} \leq s_{ij} \qquad \text{all } ij \tag{16.5b}$$

$$\sum_{j} x_{ijk\ell} = d_{i\ell} y_{k\ell} \qquad \text{all } ik\ell \tag{16.5c}$$

$$\sum_{k} y_{k\ell} = 1 \qquad \text{all } \ell \tag{16.5d}$$

$$\underline{v}_k z_k \leq \sum_{i\ell} d_{i\ell} y_{k\ell} \leq \bar{v}_k z_k \qquad \text{all } k \tag{16.5e}$$

Linear configuration constraints on y and z (16.5f)

$$x_{ijk\ell} \geq 0 \qquad \text{for all } ijk\ell$$

$$y_{k\ell} = 0 \text{ or } 1 \qquad \text{for all } k, \ell \tag{16.5g}$$

$$z_k = 0 \text{ or } 1 \quad \text{for all } k$$

The constraints (16.5b) limit the supply of commodity that can be shipped from plant j. The constraints (16.5c) and (16.5d) together state that the demand for commodity i by customer ℓ must be met and by shipment from exactly one warehouse. The constraints (16.5e) state that if warehouse site k is selected ($z_k=1$), then total storage of all commodities for all customers supplied from k must be between the lower and upper limits $\underline{v}_k$ and $\bar{v}_k$. The constraints (16.5f) are a variety of logical constraints on the zero-one decision variables such as

$$\sum_{k \in K^1} z_k \leq 1$$

implying no more than one warehouse site can be selected from a subset K^1 of the possible sites. Finally, the objective function (16.5a) consists of linear terms and fixed change terms involving the variables $y_{k\ell}$ and z_k.

For the application of Geoffrion and Graves, there were 17 different commodities, 14 plants, 45 possible warehouse sites and 121 customers. The mixed integer programming problem (16.5) consisted of 11,854 rows, 727 binding variables and 25,513 continuous variables. These large figures are somewhat misleading because the continuous part of the problem consists of a number of transportation problems with simple structure. Fortunately, it was possible to exploit these structures, and at the same time solve the mixed integer programming problem, by the use of Benders' method for mixed integer programming as shown schematically in Figure 16.1.

The integer programming subproblem (IP) involved the variable $y_{k\ell}$ and z_k and the constraints (16.5d), (16.5e), (16.5f) and the zero-one constraints in (16.5g) plus constraints approximating the objective function (16.5a) from below. The transportation models

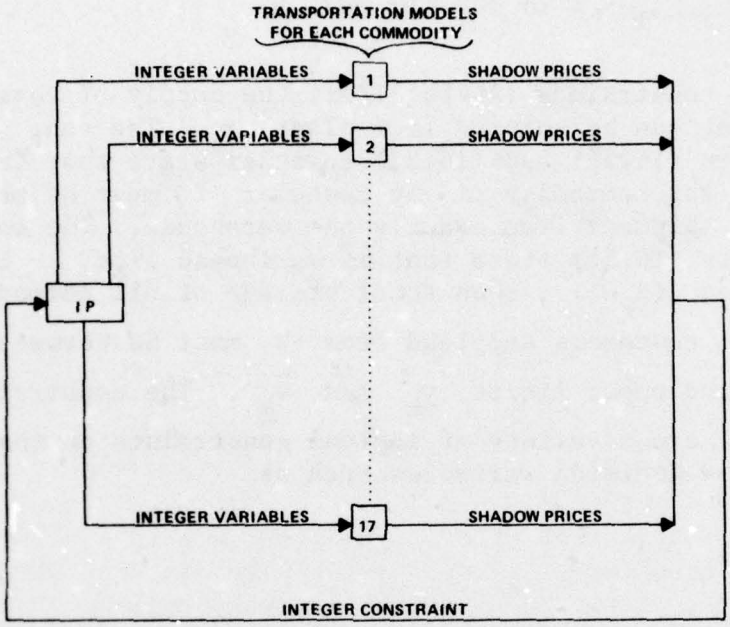


Figure 16.1 - Solution of a mixed integer programming problem.

consisted for each commodity i of (16.5b) and (16.5c) where the variables $y_{k\ell}$ were fixed at zero-one values. The objective functions consisted of the linear terms $\sum_{jk\ell} c_{ijk\ell} x_{ijk\ell}$ for each commodity i. Benders' method proceeds by alternatively solving the integer programming subproblem and the continuous transportation problem. It stops when the integer constraint derived from the transportation subproblems does not cut off the previously optimal solution to the integer programming subproblem.

As we indicated, each solution of IP produced a better lower bound to the optimal objective function value in (16.5). Moreover, each solution of the 17 transportation problems produced a feasible mixed integer programming solution to (16.5). Thus, it is possible to terminate computation before optimality is reached (or proved), and have a bound on the objective function cost loss due to nonoptimality.

Illustrative Application Three. Optimal Design of

Offshore Natural-Gas Pipeline System (Rothfarb et al. (1970)).

The previous two examples have involved continuous as well as discrete decision variables and therefore they required mixed methods of solution. Specifically, dual pricing of scarce resources was required in order to adjust the costs on discrete decision variables. By contrast, the application to be discussed here is purely discrete and requires combinatorial algorithms adapted from algorithms for simpler problems of similar type. Moreover, the complexity of the problem necessitates the use of heuristic methods because optimality is too costly to obtain.

Figure 16.2 depicts a typical design of a pipeline system connecting offshore gas fields (nodes) to an onshore separation and compressor plant. The location of the fields is assumed given and the graph of the system is always a tree (that is, one and only one path from a gas field to the plant). The pipeline system is required to carry known flow per day from each gas field according as

$$\text{flow} = K \frac{(\text{pressure change})^2}{(\text{pipe length})} (\text{pipe diameter}) \qquad (16.6)$$

where K is a proportionality constant. The variables on the right side of (16.6) are the design parameters. In addition, there are upper bound constraints on pressure due to safety and design considerations, and lower bounds due to delivery requirements at the plant. The cost of a pipeline link depends on its diameter and the depth of the water. The plant costs depend upon flow and delivered pressure.

The two main problems addressed by Rothfarb et al. were the following.

(1) The selection of minimal cost pipeline diameter given a pipeline network and delivery requirements

(2) The design of a minimal cost pipeline network given gas field locations and delivery requirements

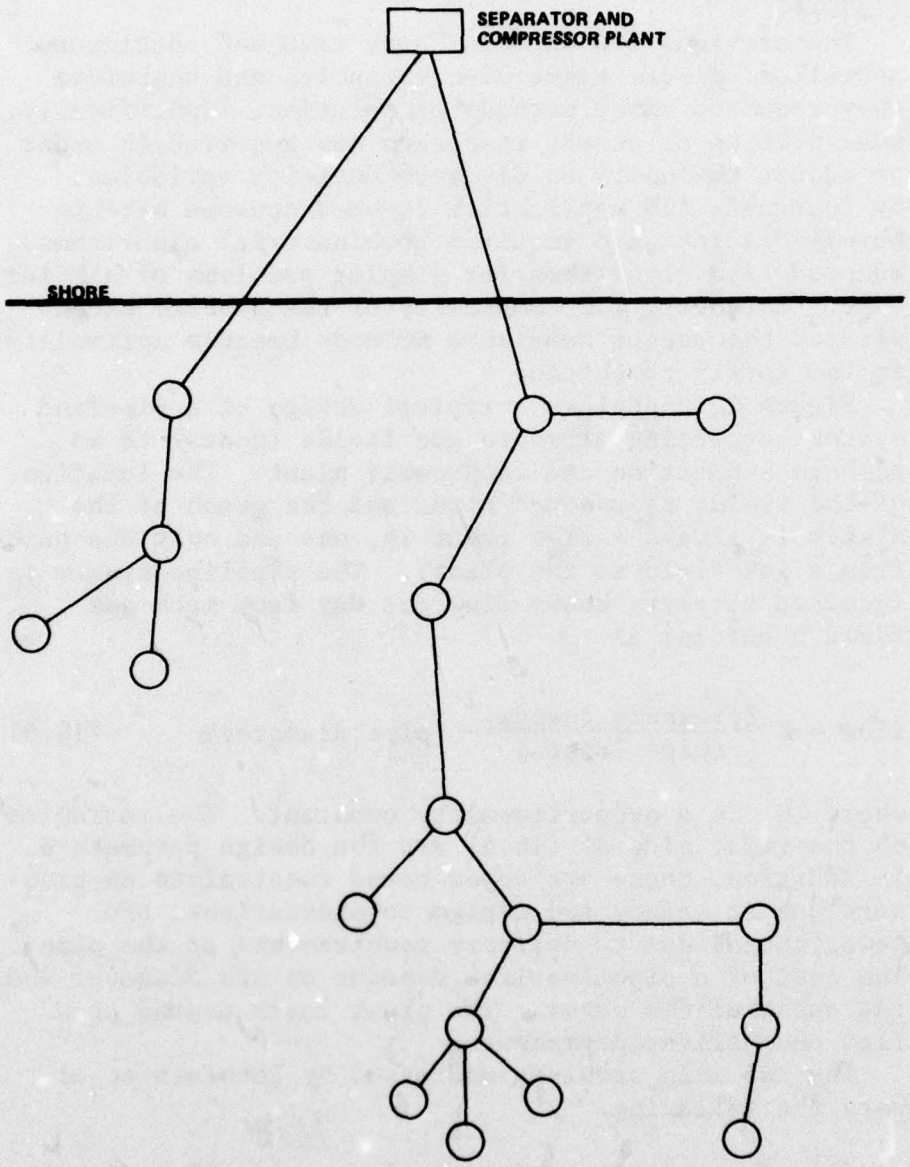


Figure 16.2 - Typical design of an offshore pipeline system.

Problem (1) is a subproblem of Problem (2).

Problem (1) was surprisingly difficult to optimize because the relation (16.6) and the pipe costs are nonlinear, the number of different pipe diameters was 7 and the number of gas fields was 20 or more. As a result, the number of design combinations was quite large and the nonlinearities made it difficult to identify dominating subsets of the combinations. Heuristic rules were developed to eliminate apparent uneconomical diameter combinations without exhaustive enumeration. The heuristics were based on looking at critical paths which are those to the ends of the trees where the flow and therefore the pipe diameters are smallest. The heuristics entailed local optimization at these ends followed by a merging of the nodes at the end into a single node with an aggregate design and flow requirements. The analysis was then repeated on the reduced network.

Problem (2) subsumes Problem (1) and required additional heuristics. First, it is known that the pipes connected directly to the plants, called arms, play an important role in determining overall cost. It is assumed either that these are given by the user, or Problem (2) must be solved for all possible combinations of arms. An automatic tree generator is used to generate a distribution of candidates for solution. The following two guidelines were used.

(i) Efficient trees have low total pipe length

(ii) Efficient trees have nearly equal flow in their arms

If the first guideline was the only criterion, then the problem of pipeline network design could be solved as a minimum spanning tree problem by a "good" algorithm.

This illustrative application is only one of many examples of network design and analysis for problems where exact optimization is difficult. An attractive possibility is to use man-machine interactive computer programs to find satisfactory designs. Such a program has been constructed by Schneider et al. (1972) to design urban transportation networks.

A class of network design problems from an entirely

different application area giving rise to optimization problems with similar mathematical structure are computer communications network design problems. A number of remote terminals are to be attached to a central computer by a communications network. The costs to be minimized are line costs plus concentrator costs for those nodes where many lines are accumulated. See Frank et al. (1971) for a discussion of models of this type.

Application Four. Routing Problems.

We have not found in the literature a single application of the routing problem illustrating many of its aspects. A simple version of this problem is the following. A trucking company must deliver a quantity q_i of a single commodity to customer i for $i = 1, \ldots, m$. The company has an unlimited number of trucks of capacity Q which can transport the commodity from the warehouse to the customers. We assume $q_i \leq Q$ for all i and orders cannot be split between two or more delivery trucks. The objective is to minimize the total distance traveled by the delivery trucks. Let $d_{ij} = d_{ji}$ denote the distance from customer i to customer j where d_{0j} is the distance from the warehouse to customer j. Figure 16.3 depicts a typical problem of this type with a solution involving four trucks.

An integer programming formulation of the problem has been given by Balinski and Quandt (1964). A generic activity a_j, called a tour, is an m-vector with components

$a_{ij} = 1$ if delivery route j visits customer i

$= 0$ otherwise

where the a_{ij} satisfy

$$\sum_{i=1}^{m} a_{ij} q_i \leq Q$$

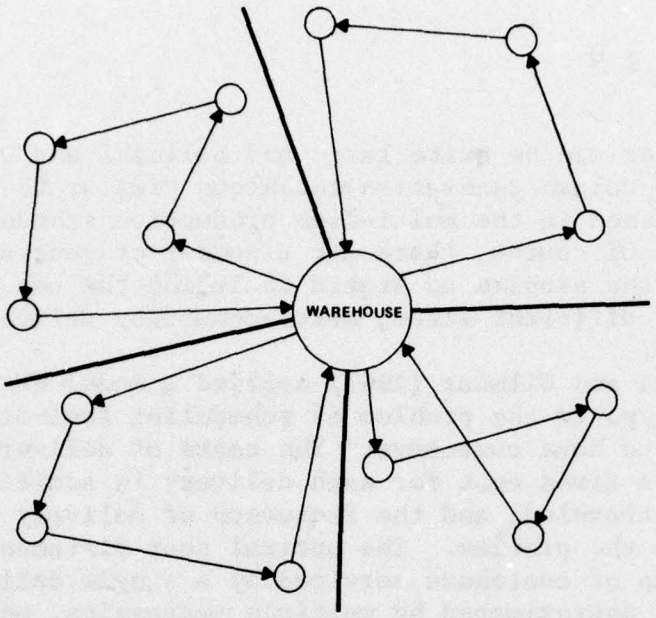


Figure 16.3 · A solution involving four trucks.

The objective function coefficient c_j associated with a_j is the shortest distance tour, starting and ending at the warehouse, of the customers visited by the activity. The calculation of c_j is a traveling salesman problem. The delivery problem is solved by solving the set partitioning problems.

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^{n} a_{ij} x_j = 1 \quad i = 1,\ldots,m \\ & x_j = 0 \text{ or } 1 , \quad j = 1,\ldots,n \end{aligned} \tag{16.7}$$

where n is the total number of tours satisfying

$$\sum_{i=1}^{n} a_{ij} q_i \leq Q$$

This number can be quite large and Balinski and Quandt suggest a column generation technique similar to the one discussed in the multi-item production scheduling example. Of course, there are a number of generalizations of the problem as stated including the use of trucks of different sizes, multi-commodity delivery, and so on.

Hausman and Gilmour (1967) applied a model of this general type to the problem of scheduling fuel-oil delivery to home customers. The costs of delivery included a fixed cost for each delivery in addition to distance traveled, and the frequency of delivery was a factor in the problem. The optimal tour distance for each group of customers serviced by a single delivery truck was approximated by multiple regression, using a few simple statistics for the group. Practical problems involving 120 customers were solved with a substantial cost reduction over hand solutions.

An important class of routing problems with the form (16.7) are the airline crew scheduling problems (for example, see Arabeyre et al. (1969) and Simpson (1969)). For these problems, the "customers" are cities and the "warehouse" is a home base for crews and planes. A route map is given with the existing flights, and their times, that must be flown between cities during a given time period, usually a few days or a week. An activity a_j corresponds to a sequence of cities connected by flights that can be flown by a crew without violating safety and union constraints. The cost c_j of such an activity consists of the bonuses, per diem and overtime payments. In practical applications of the airline crew scheduling problem, there can be more than one home base for crews, and additional constraints limiting the number of crews that can begin and end their tours at each home base.

Laderman (1966) and Lasdon (1973) have formulated and solved some large routing problems for ships on the Great Lakes. Mevert (1974) reports on a large trans-Atlantic shipping problem that has been formulated as a

problem of the type (16.7) with a number of side constraints.

16.3 Conclusions

We have tried to present applications of integer and combinatorial programming in logistics that illustrate the current state of the art of these methods and some principles to be applied to new applications. There are a number of application areas that were not mentioned, including for example, reliability (Kershenbaum and Van Slyke (1972)), decision CPM (Crowston (1970)), and the setting of traffic signals (Little (1966)). Finally, we have tried to indicate a representative rather than an exhaustive list of references. Extensive bibliographies can be found in Garfinkel and Nemhauser (1972) and Scott (1970).

References

[16.1] Arabeyre, J. P., J. Fearnley, F. C. Steiger, and W. Teather (1969). The airline crew scheduling problem: a survey. Transportation Sci. 3 140-163.

[16.2] Balinski, M. L., and R. E. Quandt (1964). On an integer program for a delivery problem. Operations Res. 12 300-304.

[16.3] Crowston, W. B. (1970). Decision CPM: network reduction and solution. Operations Res. Quart. 21 435-452.

[16.4] Edmonds, J. (1971). Matroids and the greedy algorithm. Math. Programming 1 127-136.

[16.5] Fisher, M., W. Northup, and J. Shapiro (1974). Using duality to solve discrete optimization problems: theory and computational experience. Working Paper 030-74, Operations Research Center, Massachusetts Institute of Technology.

[16.6] Frank, H., I. T. Frisch, R. Van Slyke, and W. S. Chou (1971). Optimal design of centralized computer networks. Networks 1 43-57.

[16.7] Garfinkel, R. S., and G. L. Nemhauser (1972).

Integer Programming. Wiley.

[16.8] Geoffrion, A., and G. Graves (1973). Multicommodity distribution system design by Benders' decomposition. Working Paper 209, Western Management Science Institute, University of California, Los Angeles.

[16.9] Geoffrion, A., and R. Marsten (1972). Integer programming algorithms: a framework and state of the art survey. Management Sci. 18 465-491.

[16.10] Glover, F., D. Karney, D. Klingman, and A. Napier (1974). A computational study on start procedures, basis change criteria, and solution algorithms for transportation problems. Management Sci. 20 793-813.

[16.11] Gorry, G. A., W. Northup, and J. Shapiro (1973). Computational experience with a group theoretic integer programming algorithm. Math. Programming 4 171-192.

[16.12] Hausman, W. H., and P. Gilmour (1967). A multi-period truck delivery problem. Transportation Research 1 349-357.

[16.13] Held, M., and R. M. Karp (1970). The traveling salesman problem and minimum spanning trees. Operations Res. 18 1138-1162.

[16.14] Held, M., and R. M. Karp (1971). The traveling salesman problem and minimum spanning trees: Part II. Math. Programming 1 6-25.

[16.15] Himmelblau, D. M. (ed.) (1973). Decomposition of Large-Scale Problems. North-Holland.

[16.16] Karp, R. M. (1972). Reducibilities among combinatorial problems. Computer Science Report 3, University of California, Berkeley.

[16.17] Kershenbaum, A., and R. Van Slyke (1973). Recursive analysis of network reliability. Networks

3 81-94.

[16.18] Laderman, J., L. Gleiberman, and J. F. Egan (1966). Vessel allocation by linear programming. Naval Res. Logist. Quart. 13 315-320.

[16.19] Lasdon, L. S. (1970). Optimization Theory for Large Systems. MacMillan.

[16.20] Lasdon, L. S. (1973). Decomposition of a ship routing problem. in Himmelblau (1973) 235-240.

[16.21] Ladson, L. S., and R. Terjung (1971). An efficient algorithm for multi-item scheduling. Operations Res. 19 946-969.

[16.22] Lea, A. C. (1973). Location-allocation systems: an annotated bibliography. Discussion Paper 13, Department of Geography, University of Toronto.

[16.23] Little, J. D. C. (1966). The synchronization of traffic signals by mixed integer linear programming. Operations Res. 14 568-594.

[16.24] Magnanti, T. L., J. Shapiro, and M. H. Wagner (1973). Generalized linear programming solves the dual. Working Paper 019-73, Operations Research Center, Massachusetts Institute of Technology.

[16.25] Mevert, P. (1974). Personal Communication.

[16.26] Müller-Merbach, H. (1973). Switching between bill of material processing and the simplex method in certain linear large-scale industrial optimization problems. in Himmelblau (1973) 189-200.

[16.27] Rothfarb, B., H. Frank, D. M. Rosenbaum, K. Steiglitz, and D. J. Kleitman (1970). Optimal design of offshore natural-gas pipeline systems. Operations Res. 18 992-1020.

[16.28] Schneider, A., G. Symons, and A. Goldman (1972). Planning transportation terminal systems in urban regions. A man-machine interactive problem-

solving approach. Transportation Research 6 257-273.

[16.29] Scott, A. J. (1971). Combinatorial Programming, Spatial Analysis and Planning. Methuen.

[16.30] Simpson, R. W. (1969). Scheduling and routing models for airline systems. Report R68-3, Flight Transportation Laboratory, Massachusetts Institute of Technology.

[16.31] Wagner, H. M. (1969). Principles of Operations Research. Prentice-Hall.

NAME INDEX*

Adolphson, D. 209, 218
Agin, N. 133, 201, 278
Agnew, R. A. 359
Ahrens, J. H. 368
Allen, S. G. 278, 377
Almogy, Y. 372
Alway, G. G. 358
Anthony, R. N. 63-66, 89
Antosiewicz, H. A. 385
Arabeyre, J. P. 432, 433
Armstrong, R. J. 75, 85, 89
Arrow, K. J. 70, 90, 255, 279, 285, 286, 290, 331, 370
Arthanari, T. S. 374
Ashour, S. 133, 201
Astrachan, M. 55, 224, 243
Atkins, R. J. 70, 90
Aumann, R. J. 385

Balachandran, V. 388, 397, 399, 401, 402, 412, 413
Balas, E. 374
Balinski, M. L. 373, 430, 432, 433
Balintfy, J. L. 279
Ballou, R. H. 71, 90
Barankin, E. W. 279
Barlow, R. E. 241, 243, 339, 340, 346
Baron, D. P. 369
Barr, R. S. 356n
Bartholdi, J. J. 214, 220
Bartlett, T. E. 373, 374, 377
Barton, L. G. 167, 205
Bather, J. A. 252, 279
Baumol, W. J. 373
Bayes, T. See Subject Index
Bazaraa, M. S. 288
Beale, E. M. L. 133, 201, 358, 359, 363, 368
Beckmann, M. 279
Beesack, P. R. 279
Beged-Dov, A. G. 377
Bellman, R. 279, 370, 377, 378
Bellmore, M. 363, 378, 383
Bell, C. F. 55
Benders, J. F. 184, 187, 191, 199, 201, 425, 426
Bennington, G. 378, 383
Bent, D. H. 79, 95
Berman, M. B. 55
Berry, S. D. 367
Bessler, S. A. 279
Blachman, N. M. 384
Blackett, D. W. 360
Blackwell, D. 359, 360
Blitz, M. 378
Blumenthal, S. C. 71,90
Bohlinger, J. D. 220
Boldyreff, A. W. 363
Boodman, D. M. 70, 94
Bowersox, D. J. 71,90
Bowman Jr., V. J. 366
Box, G. E. P. 70, 90
Boylan, E. S. 280
Bracken, J. 369, 378, 379, 382
Braithwaite, R. B. 360
Brandenburg, R. G. 386
Bratley, P. 374
Breining, P. 366
Bres III, E. S. 351 (Chapter 14)

*Underlined page numbers denote locations of complete references.

Breuer, M. A. 378
Breu, R. 157, 201
Briggs, F. E. A. 363
Brooks, G. H. 70, 94
Brooks, R. B. S. 55
Brown, B. B. 55, 56, 57, 58
Brown Jr., G. F. 280
Brown, R. G. 70, 90
Bryson, M. 337, 346
Buchanan, A. L. 56
Buffa, E. S. 70, 90
Burdet, C. A. 157, 201
Burnham, P. R. 378
Burt Jr., J. M. 374
Buzacott, J. A. 374

Cahn, A. S. 224, 243
Campbell, H. S. 56
Carter, G. M. 206, 218
Cass, D. 355n, 356n
Chaiken, J. M. 206, 218
Chambers, J. C. 70, 90
Chandrasekaran, R. 381
Chan, A. W. 210, 218
Charnes, A. 351 (Chapter 14), 352n, 353n, 355n, 356n, 357n, 360, 363, 366, 369, 370, 371, 373, 379, 390, 406, 409, 413
Chattopadhyay, R. 360
Chebyshev, F. L. 406
Cheney, L. K. 379
Chern, H. C. 280
Chernoff, H. 331
Choe, U. C. 260, 273, 291
Chou, W. S. 430, 433
Christofides, N. 70, 71, 91
Clark, A. J., 56, 71, 90, 260, 280, 294
Clark, C. E. 167, 201
Cobb, C. W. 106
Cohen, I. K. 56
Cohen, J. J. 69, 70, 91
Cohen, N. D. 360
Connors, M. M. 60, 81, 86, 91, 280
Control Data Corporation, 79, 90
Conway, R. W. 56, 71, 91
Cooper, C. R. 382
Cooper, L. 373
Cooper, W. W. 351 (Chapter 14), 352n, 353n, 355n, 356n, 357n, 363, 366, 369, 370, 371, 390, 406, 409, 413
Coray, C. 60, 81, 86, 91
Corcoran, T. M. 280
Cremeans, J. E. 374
Crowston, W. B. 208, 220, 280, 368, 433
Cuccaro, C. J. 60, 81, 86, 91
Curry, G. L. 284

D'Esopo, D. A. 371, 379, 386
Dade, M. A. 56
Daniel, K. 281
Danskin, J. M. 360
Dantzig, G. B. 363, 366, 369, 372, 373, 379, 383, 390, 411, 413, 422
Daubin, S. C. 379
Davis, E. W. 133, 201
Davis, M. 360
Davis, P. S. 379
Day, J. E. 71, 91, 374
Dean, B. V. 289
Deemer, R. L. 268n, 291
De Groot, M. 331
Delfausse, J. J. 56

Dellinger, D. C. 367, 379
Denardo, E. V. 56
Denicoff, M. 224, 227, 243, 244, 246, 281
Denzler, D. R. 373
Derman, C. 243, 244, 384, 385
Dix, L. P. 164, 202
Dixon, H. L. 386
Dobbie, J. M. 385
Doig, A. G. 160, 164, 204
Donis, J. N. 379
Douglas, P. H. 106
Drake, W. W. 56
Draper, N. R. 70, 91
Drebes, C. 373
Dreyfus, S. 378
Drezner, S. M. 56
Dutta, S. K. 374
Dvoretzky, A. 281
Dwyer, P. S. 364, 373

Eastman, W. L. 162, 163, 200, 201
Edmonds, J. 418, 433
Edwards, W. 57
Egan, J. F. 377, 381, 432, 435
Ehrenfeld, S. 241, 244
Eilon, S. 70, 71, 91, 281
Eisenberger, I. 237, 238, 245
Eisenman, R. L. 360
Eklof, W. D. 363
Elmaghraby, S. E. 70, 91, 133 (Chapter 8), 145, 162-164, 167, 191, 192, 200, 202, 205, 218, 374, 415
Elmaleh, J. 281
Elshafei, A. N. 133 (Chapter 8), 183, 188, 202, 206, 218
Emmons, H. 191, 202
Enslow, P. H. 385
Enzer, H. 367
Ericson, W. A. 371
Esary, J. D. 340, 342, 347
Eselson, L. 364
Evans, G. C. 131
Evans, G. W. 379
Evans, J. P. 371, 374
Evans, R. V. 281
Even, S. 162, 163, 201

Faaland, B. H. 202, 203
Falk, J. E. 133, 167, 176, 203
Färe, R. 102, 103, 115, 117, 132
Fearnley, J. 432, 433
Feeney, G. J. 281
Feller, W. 296, 317, 331
Fennell, J. 224, 227, 230, 234, 243, 244, 247, 262, 281, 294
Fenske, W. J. 242, 243, 244, 247
Ferguson, R. O. 357n
Finnegan, F. 57
Firstman, S. I. 379
Fisher, M. 423, 433
Fisher, R. R. 56
Fitzpatrick, G. R. 379
Flood, M. M. 386
Florian, M. 374
Ford Jr., L. R. 363, 364, 383, 399, 409, 413
Forrest, J. J. H. 154-157, 159, 203
Francis, R. L. 70, 91,

(Chapter 9) 207, 208, 210, 213, 2..4, 218, 219, 220
Frank, H. 427, 430, 433, 435
Frank, M. 368
Frisch, H. 367
Frisch, I. T. 430, 433
Fromowitz, S. 282
Fukuda, Y. 282
Fulkerson, D. R. 167, 170, 173, 176, 203, 363, 364, 379, 383, 399, 409, 410, 413

Gabbay, H. 69, 71, 92
Gaddis, W. D. 21 (Chapter 3)
Gaddum, J. W. 364
Gainen, L. 380
Gale, D. 102, 106, 131, 367, 371
Gallagher, D. J. 282
Galler, B. A. 364
Galliher, H. P. 282
Garfinkel, R. S. 157, 203, 364, 405, 413, 418, 433
Garstka, S. J. 402, 413
Gaschutz, G. K. 368
Gassner, B. J. 364
Gass, S. I. 358, 380
Gauss, C. F. 328, 330
Gaver, D. P. 235, 244, 282, 296 (Chapter 12), 326, 331, 332, 374, 390, 391, 413
Geisler, M. A. 35 (Chapter 5), 56, 57, 58, 332
Geoffrion, A. M. 70, 79, 92, 150, 197, 203, 206, 219, 398, 413, 417, 423, 425, 434
Ghare, P. M. 384
Giffler, B. 375
Gilbert, J. C. 380
Gilford, D. M. 281
Gillen, C. A. 55
Gilmour, P. 432, 434
Ginsberg, A. S. 57, 58
Gittens, J. C. 288
Gleaves, V. B. 375
Gleiberman, L. 377, 381, 432, 435
Glicksberg, I. 279
Glicksman, S. 364
Glover, F. 150, 203, 356n, 363, 384, 409, 413, 418, 434
Gnedenko, B. V. 298, 332
Goldman, A. 429, 435
Goldschen, D. Y. 241, 244
Goldstein, J. M. 70, 91, 206, 219
Golovin, J. 69-71, 75, 76, 92, 93
Gomory, R. E. 365, 366
Gorry, G. A. 417, 434
Gould, F. J. 374
Gourary, M. H. 386
Goyal, S. K. 282
Graves, G. W. 197, 203, 358, 423, 425, 434
Green, W. K. 60, 81, 86, 91
Greenberg, H. 204, 282, 358, 368
Griesmer, J. H. 360, 361
Grinold, R. C. 331
Groff, G. K. 70, 92
Gross, D. 70, 92, 248 (Chapter 11), 260, 274, 282, 283, 289, 331, 380
Gross, O. 279

Haber, S. E. 224-226, 233, 243, 244, 281, 283
Hadley, G. 70, 92, 227, 244, 283, 295, 358, 365, 380
Haines, A. L. 337, 343, 345, 346, 347
Haley, K. B. 206, 218
Hale, J. K. 361
A-Hameed, M. S. 341, 347
Hammer, P. L. 364, 389, 413, 414
Hanan, M. 208, 219
Hanssmann, F. 70, 92, 283, 294, 381
Harary, F. 386
Harris, C. M. 240, 244, 282, 284, 331
Harris, F. 248, 254, 284
Harris, M. Y. 358
Harris, T. 255, 279
Hartfiel, D. J. 284
Hartman, J. K. 358
Hartung, P. H. 284
Hatch, R. S. 409, 414
Hausman, W. H. 280, 284, 432, 434
Hax, A. C. 59 (Chapter 6), 69-71, 75, 76, 82, 85, 89, 92, 93
Hayes, R. H. 284
Haythorn, W. W. 57
Hegerick, R. L. 204
Held, M. 165, 200, 204, 208, 219, 418, 434
Hempley, R. B. 359
Hershkowitz, M. 361, 386
Heselden, G. P. M. 359
Hicks, J. R. 106
Hildreth, C. 368
Hillier, F. S. 203
Himmelblau, D. M. 434
Hirsh, W. M. 372, 373
Hirst, J. P. H. 154-157 159, 203
Hitchcock, D. F. 380
Hitchcock, F. L. 363
Hixon, O. M. 56
Hochstaedter, D. 284
Hoekstra, D. 268n
Hoffman, A. J. 364, 366, 371
Holladay, J. 364
Hollander, M. 344, 347
Holt, C. C. 70, 93, 285
Honig, D. P. 380
Horowitz, J. L. 133, 167, 176, 203
Hottenstein, M. P. 71, 91, 374
Houston, B. F. 380
Howard, G. T. 380
Howard, R. A. 295, 303, 332
Ho, C. M. 284, 291
Huffman, R. A. 380
Hull, C. H. 79, 95
Hunter, L. C. 237, 241, 243, 245
Hunt, J. 285
Hunt, R. B. 380
Hurter, A. P. 285
Hurwicz, L. 370
Hu, T. C. 209, 218, 375, 384

IBM Corporation 60, 75, 78-81, 93, 94
Iglehart, D. L. 253, 258n, 285, 286, 294, 327, 330, 331, 332
Ignall, E. 161, 204, 206, 218, 286
Isaacs, J. M. 162, 163, 201

Isaacs, R. 386
Isbell, J. R. 361, 385

Jackson, J. R. 375, 386
Jackson, J. T. R. 382
Jacobs, W. 358, 364
Jarvis, J. J. 384
Jelmert, A. E. 250n
Jenkins, G. M. 70, 90
Jewell, W. S. 380
Johansen, L. 102-106, 109, 110, 131
Johnson, E. L. 286
Johnson, L. 364
Johnson, S. M. 373, 375
Joksch, H. C. 371

Kaminsky, F. C. 285
Kamins, M. 242, 245
Kampe, W. R. 280
Kan, A. H. G. R. 133, 204
Kander, Z. 242, 245
Kantorovich, L. V. 363
Kaplan, A. J. 286
Kaplan, R. J. 57
Kaplan, R. S. 286
Karel, C. 148, 181, 200, 204
Karlin, S. 70, 90, 255, 285, 286, 287, 290, 331, 370, 371, 381
Karmarkar, U. S. 69, 71, 76, 94
Karney, D. 409, 413, 418, 434
Karp, R. M. 165, 200, 204, 208, 219, 418, 434
Karreman, H. F. 381
Karr, H. W. 57, 287
Karush, W. 287, 371, 381
Kasegai, T. 287
Kasugai, H. 287
Keilson, J. 331, 332
Kelley Jr., J. E. 358, 381
Kershenbaum, A. 433, 434
Keswani, A. K. 288
Kiefer, J. 281
Kiviat, P. J. 56
Klein, M. 375, 384, 385
Kleitman, D. J. 427, 435
Klingman, D. 356n, 363, 409, 413, 414, 418, 434
Kolesar, P. J. 381
Koopmans, T. C. 102, 106, 131, 363
Kornet Jr., F. 11 (Chapter 2), 21
Kortanek, K. O. 370, 375
Kruskal, J. B. 385
Kuhn, H. W. 363, 365, 373, 402, 414
Kurtzberg, J. M. 208, 219

Lagrange, J. L. 421, 422
Laderman, J. 288, 377, 381, 432, 435
Lagemann, J. J. 365
Lalchandani, A. 285
Land, A. H. 160, 164, 204
von Lanzenauer, C. H. 376
Laplace, P. S. 324, 325
Larson, R. C. 206, 219
Lasdon, L. S. 358, 419, 421, 422, 432, 435
Lawler, E. L. 133, 204
Lea, A. C. 206, 219, 423, 435
Learner, D. B. 355n
LeCam, L. M. 332, 347
Lefkowitz, B. 379, 386

Lehoczky, J. P. 332
Lemke, C. E. 358, 371
Lemoine, A. J. 330, 332
Leontief, W. W. 102, 106, 113, 114, 131, 366
Levin, O. 372
Levitan, R. E. 360
Levy, F. K. 375
Levy, J. 288
Lewis, P. A. W. 332
Lindeman, A. R. 382
Littauer, S. B. 288
Little, J. D. C. 148, 181, 200, 204, 433, 435
Lloyd, R. M. 280
Lomnicki, Z. A. 160, 162, 204
Longhill, J. D. 378
Lorden, G. 237, 238, 245
Love, R. F. 288, 381, 383
Low, D. W. 60, 81, 86, 91
Lubore, S. H. 378, 383, 384
Lucas, W. F. 362
Luckew, R. S. 332
Lundh, B. 293
Lu, J. Y. 55

McCloskey, J. F. 381
McCormick, A. 262
McGlothlin, W. H. 230, 245
Machol, R. 352n
McIver, D. W. 57
McMasters, A. W. 384
McNamara, R. S. 15
McNeill, D. R. 327, 332, 333
MacQueen, J. B. 380
McShane, R. E. 381
Madow, W. G. 381
Magee, J. F. 70, 71, 94, 295
Magnanti, T. L. 423, 435
Maher, M. J. 288
Malik, H. J. 371
Mallik, A. K. 164, 202
Mangasarian, O. L. 371
Manne, A. S. 288, 367
Marchal, W. G. 240, 244
Marchi, E. 361
Markland, R. E. 227-229, 245
Markov, A. A. See Subject Index
Markowitz, H. M. 60, 81, 86, 91, 364, 368
Marks, B. G. 56
Marlow, W. H. 224, 243, 281, 361
Marshak, J. 255, 279
Marshall, A. W. 338, 340, 342, 343, 347
Marshall, K. T. 331, 359
Marsten, R. 417, 434
Martin Jr., J. I. 367
Martos, B. 371
Maschler, M. 360, 361
Maxwell, W. L. 57, 71, 91, 375
Mazumdar, M. 235, 244, 245
Meal, H. C. 75, 76, 85, 93
Mellon, W. G. 386
Mendolia, A. I. 3 (Chapter 1), 11
Menon, V. V. 384
Merrill, K. 58
Mevert, P. 432, 435
Meyer, M. 372
Midler, J. L. 369
Miller, D. W. 295
Miller, L. W. 57, 71, 91
Miller, M. 371

Miller, M. H. 379
Miller, R. E. 220
Mitten, L. G. 133, 135, 137-142, 177-179, 195, 204, 205
Mize, J. H. 70, 94
Modigliani, F. 70, 93, 285
Moglewer, S. 361
Mond, B. 361
Montgomery, D. C. 288, 384
Morey, R. C. 285, 286, 288
Morgan, R. W. 288
Morgenstern, O. 359, 367
Morrill, J. E. 361
Morse, P. M. 282, 288
Morton, T. E. 262, 289
Muckstadt, J. A. 260, 289
Mueller, W. A. 248, 254, 293
Mukhopadhyay, A. C. 374
Müller-Merbach, H. 423, 435
Mullich, S. K. 70, 90
Murdick, R. G. 71, 94
Murty, K. G. 148, 181, 200, 204
Mustin, T. M. 384
Muth, J. F. 70, 71, 92, 93, 94, 285
Muth, R. 279

Naddor, E. 70, 95, 262, 295
Nahmias, S. 250n
Nair, K. P. K. 381
Naor, P. 242, 245
Napier, A. 409, 413, 414, 418, 434
Nauta, F. 409, 414
Neghabat, F. 219
Nelson, C. R. 70, 95
Nemhauser, G. L. 157, 203, 359, 363, 366, 380, 418, 433
von Neumann, J. 359, 362, 366
Neuts, M. 289
Newman, D. J. 217, 219
Neyman, J. 332, 347
Nie, N. 79, 95
Niehaus, R. J. 355n, 356n, 406, 413
Noble, S. B. 381, 386
Northup, W. 417, 423, 433, 434

O'Brien, M. J. 379
O'Hagan, M. 293
O'Neill, R. R. 375, 381, 385, 386
Oliver, R. M. 331
Oral, M. 289
Orgler, Y. 397, 414
Ornstein, L. S. 327, 330
Owen, G. 361

Page, E. S. 238-240 246
Papineau, R. L. 214, 219, 220
Pareto, V. 353, 398
Park, S. 162, 191, 192, 202
Payne, C. 361
Pegels, C. C. 250n
Peleg, B. 361, 362
Pennington, A. W. 382
Pentico, D. 262
Perlas, M. 332, 374

Peterson, R. 284
Pfanzagl, J. 386
Pierce, J. F. 208, 220, 368
Pierce, M. B. 409, 414
Pierskalla, W. P. 250n
Pinkus, C. E. 260, 274, 289
Poisson, S. D. See Subject Index
Pollack, M. 382
Pollack, S. 220
Pollack, S. M. 379
Porteus, E. L. 289
Posner, M. J. 289
Prager, W. 365, 375
Pratt, V. R. 208, 210, 220
Prichard, J. W. 274n
Proschan, F. 237, 241, 243, 245, 334 (Chapter 13), 338-344, 346, 347
Pruitt, W. E. 362, 381
Pruzan, P. M. 382

Quandt, R. E. 367, 430, 432, 433
Quesnay, F. 102
Quraishi, M. N. 133, 201

Rabinovitch, A. 242, 245
Radner, R. 230, 245
Randolph, P. H. 372
Rao, M. R. 364, 405, 413
Ratliff, H. D. 384
Rau, J. G. 382
Ravindran, A. 289
Ray, T. L. 379
Rech, P. 167, 205
Reisman, A. 289
Rice, E. W. 382
Richardson, M. 386
Rigby, F. D. 365
Ritter, K. 372
Roach, C. 250n
Roberts, D. M. 290
Robillard, P. 366, 374
Robinson, A. I. 57
Roeloffs, R. 385
Rogers, W. F. 280
Rosenbaum, D. M. 427, 435
Rose, M. 290
Rosholdt, E. F. 380
Ross, H. 331, 332
Ross, J. E. 71, 94
Ross, S. 340, 347
Rothfarb, B. 427, 435

Saaty, T. 358, 372
Sakaguchi, M. 362
Salkin, H. M. 366
Salvador, M. S. 289
Salveson, M. 387
Saposnik, R. 382
Scarf, H. 70, 90, 230-232, 246, 253, 256, 260, 280, 285, 287, 290, 292, 294, 331
Schaak, J. P. 260, 290
Schach, S. 327, 332
Scherer, F. M. 382
Schlaifer, R. 70, 95
Schneider, A. 429, 435
Schrady, D. A. 70, 92, 248 (Chapter 11), 260, 273, 291
Schrage, L. 161, 204
Schroeder, R. G. 360
Schumpeter, J. A. 101, 131
Schwartz, A. N. 382

Scott, A. J. 70, 95, 433, 436
Scott, E. L. 347
Segel, F. W. 224, 243, 281
Serfling, R. 347
Service in Informatics and Analysis, Ltd. 79, 95
Shapiro, J. F. 416 (Chapter 16), 417, 423, 433, 434, 435
Shapley, L. S. 359, 362, 370, 372, 383, 384
Sheler, J. A. 382
Shelly, M. W. 281
Shephard, R. W. 99 (Chapter 7), 102, 103, 115, 117-120, 124, 126, 131 132
Sherashian, R. 200, 204
Sherbrooke, C. C. 55, 57, 260, 262, 281, 291
Shriver, R. H. 70, 90
Shubik, M. 360, 361, 362
Shulman, H. L. 57
Sicilia, G. T. 384
Siddiqui, M. 337, 346
Sides, R. S. 284
Silver, E. A. 70, 95, 260, 274n, 290, 291
Simmons, D. M. 209, 220, 291
Simmons, K. W. 378
Simon, H. A. 70, 93
Simon, R. M. 291
Simond, M. 282
Simonnard, M. A. 358, 365
Simpson Jr., K. F. 291
Simpson, R. W. 432, 436
Singpurwalla, N. D. 240, 241, 244, 337, 343, 345, 346, 347
Sitgreaves, R. 224-226, 233, 244, 283
Sivazlian, B. D. 292
Smith, D. D. 70, 90
Smith, H. 70, 91
Smith, J. W. 382
Smith, M. W. 385
Smith, R. A. 374
Smith, T. C. 55, 57
Smith, V. L. 382
Smith, W. E. 375
Snavely, W. W. 21, 25 (Chapter 4)
Sodaro, D. 375
Sokolowsky, D. 364
Soland, R. M. 176, 203, 260, 274, 289, 369, 380
Solomon, Henry 224-227, 243, 244, 246, 281, 283, 381
Soriano, A. 282, 283
Soyster, A. L. 375
Soyster, H. R. 382
Spinner, A. H. 376
Spivey, W. A. 367
Srinivasan, V. 164, 205, 356n, 376, 388, 389, 391, 392, 395-399, 402, 403, 405, 406, 409-411, 414, 415
Stanley, E. D. 380
Starr, M. K. 70, 95, 295
Stedry, A. C. 386, 406, 413
Steger, W. A. 57
Steiger, F. C. 432, 433
Steiglitz, K. 427, 435
Steinberg, D. I. 373
Stevenson, K. A. 206, 219

Stone, C. 327
Stutz, J. 414
Suppes, P. 290
Suurballe, J. W. 359
Suzuki, G. 382, 383
Sweat, C. W. 362
Sweeney, D. W. 148, 181, 200, 204
Sweetland, A. 57
Swinson, G. E. 372
Symonds, G. H. 369
Symons, G. 429, 435
Szwarc, W. 365, 376, 382, 388, 405, 406, 410, 415

Tamura, H. 367
Tan, F. K. 292
Taubert, W. H. 70, 90
Taylor, R. J. 372
Teather, W. 432, 433
Terjung, R. 419, 421, 422, 435
Thatcher, J. W. 220
Theil, H. 70, 95
Thomas, L. J. 284
Thompson, G. L. 71, 94, 164, 205, 356n, 362, 367, 375, 376, 388 (Chapter 15), 389-392, 395-401, 402, 405, 406, 409-411, 412, 413, 414, 415
Thompson, P. M. 359
Thompson, S. P. 362, 372
Thrall, R. M. 362
Tijms, H. C. 292
Tomlin, J. A. 154-157, 159, 203
Topkis, D. 292
Tsou, C. A. 177-179, 195, 205
Tucker, A. W. 402
Turner, W. C. 384
Tyndall, G. R. 374

Uhlenbeck. G. E. 327, 330
Univac Division 79, 95
Uzawa, H. 370
van de Panne, C. 368, 369
Van Horn, R. L. 56, 58
Van Slyke, R. 430, 433, 434
Varley, T. C. 378
Vassian, H. J. 292
Vazsonyi, A. 381
Veinott Jr., A. F. 70, 95, 229, 246, 253, 279, 286, 292, 293, 294, 331
Vergin, R. C. 383
Verhulst, M. 362
von Lanzenauer See Lanzenauer
von Neumann See Neumann

Waggener, H. A. 383
Wagner, H. M. 70, 88, 95, 96, 262, 293, 359, 376, 383, 418, 436
Wagner, M. H. 423, 435
Walsh, J. E. 385
Warburton, A. R. 135, 137, 205
Wardrop, J. G. 353n
Ware, W. H. 57
Watson-Gandy, D. T. 70, 71, 91
Weibull, W. See Subject Index
Weinstock, J. K. 375
Weiss, L. 288
Wentling, L. G. 379
Wesolowsky, G. O. 383

Wheeler, A. C. 293
Whinston, A. 368, 369, 371, 372
Whinston, V. 371
White, C. R. 70, 94
White, J. A. 70, 91, 206, 208, 213, 219
Whitin, T. M. 70, 92, 227, 244, 283, 293, 295, 380, 383
Whiton, J. C. 379, 383
Wicke, H. H. 361
Wiener, N. 252
Wiest, J. D. 375
Williams, J. F. 293
Wilson, E. B. 387
Wilson, R. H. 248, 254, 268, 293
Wolfe, P. 368, 422
Wolfowitz, J. 281
Wollmer, R. D. 369, 384
Wong, Y. K. 367
Wood, D. E. 133, 204
Wortham, A. W. 387
Wright, G. 293
Wurtele, Z. S. 367

Yansouni, B. 289
Yaspan, A. 293
Yasuda, Y. 362
Yechiali, U. 370
Youngs, J. W. T. 58, 293

Zabel, E. 294
Zachrisson, L. E. 362
Zacks, S. 223 (Chapter 10), 226, 230, 232-234, 236, 237, 239, 240, 242, 244, 246, 247, 262, 294
Zangwill, W. I. 280
Zehna, P. W. 279
Ziffer, A. J. 362
Zimmer, W. J. 226, 247
Zionts, S. 372
Zwart, P. B. 372

SUBJECT INDEX*

AIR FORCE
Logistics, 25 (Chapter 4), 35 (Chapter 5)
Programs and systems, 26, 27, 38, 41, 46
Specialty codes, 42
System 66-1, 41, 46

ALLOWANCE LIST TEST PROGRAM, 227

APPLICATIONS
Airline crew scheduling, 432
Assembly line balancing, 200
Bottleneck personnel, 407
Cash management, 397, 403
Computer communications network, 430
Decision critical path, 433
Distribution, 197, 423, 430
Economic equilibrium, 366, 367
Inventory, 229, 237, 296, 303
Location-allocation, 423
Machine loading, 399
Maintenance, 306
Manpower, 306, 314, 355, 356
Mathematical programming, 376 (Bibliography)
Medical, 314, 346
Miscellaneous, 385 (Bibliography)
Network design, 429
Offshore pipeline, 427
Optimal job assignment in computer network, 402
Personnel assignment, 406
Production, 241, 423
Production and inventory, 419
Production-inventory-distribution, 75-77
Project "crashing", 166
Readiness, 235
Reliability, 241, 433
Repair, 322
Routing, 430-433
Scheduling, 145, 160, 162, 163, 177, 188, 192, 195, 200, 432
Shipping perishable goods, 406
Ship routing on Great Lakes, 432
Supply, 229, 237, 296, 303
Surveillance, 241
Traffic signals, 433
Transatlantic shipping, 432
Warehouse and distribution, 423

ASSIGNMENT PROBLEM, 164, 165, 200, 363 (Bibliography), 407-409
Bottleneck, 406
Computational times, 399
Quadratic 208, 368 (Bibliography)

AVIATION SYSTEMS COMMAND,

*Titles of references are not included.

AVIATION SYSTEMS COMMAND, (continued), 17

BAYESIAN METHODS, 226, 230-234, 237, 239, 240, 242, 243, 254, 262, 320

BENDERS' METHOD, 184, 187, 199, 425, 426

BIBLIOGRAPHIES
Assignment problems, 363
Economics, 366
Fixed charge problems, 372
Game theory, 359
Integer programming, 365
Inventory, 278
Linear programming, 357
Logistics functions, 70
Mathematical programming applications, 376
Mathematical programming theory, 370
Miscellaneous logistics, 385
Network theory, 383
Probabilistic programming, 369
Quadratic and quadratic assignment problems, 368
Rand Corporation logistics, 55
Search and surveillance problems, 384
Sequencing and scheduling problems, 373
Software packages, 79
Transportation and assignment problems, 363

BIRTHDAY PROBLEM, 296 (Section 12.2)

BRANCH AND BOUND, 133 (Chapter 8), 208, 209, 397, 423
Dicta for, 143, 149, 152, 164-166, 176, 181, 188, 191, 194, 197, 200

BUREAU OF NAVAL PERSONNEL, 355

CAPACITY PLANNING, 64

COHERENT SYSTEMS, 340

CONTIGUOUS BINARY SWITCH, 178, 179, 195

CONTINUOUS REVIEW INVENTORY POLICY, 251, 256, 257, 265, 268-270, 273, 274

CONTROL CHARTS, 238

COSAL (COORDINATED SHIPBOARD ALLOWANCE LIST), 270

COST AND REVENUE FUNCTIONS, 122 (Section 7.5)

COST BENEFIT ANALYSIS, 52, 126, 129

COST EFFECTIVENESS ANALYSIS, 16, 53

COST RETURN ANALYSIS, 126, 128

CRITICAL PATH METHODS, 373

CUTS, 181, 182, 184, 188, 191, 195

DANTZIG-WOLFE DECOMPOSITION, 422

DATA ACQUISITION, 27, 36-38, 42, 46 (Section 5.5), 224, 273, 274

DATA VARIABILITY
Exploitation of, 48 (Section 5.6)

DECOMPOSITION METHODS, 181, 195, 354, 419, 422, 423

DEMAND FORECASTING AND PREDICTION, 40, 223 (Section 10.2), 233, 252, 254, 262, 265, 270-274

DISCRETE DISTRIBUTION, 236

DISCRETE PROGRAMMING, 416 (Chapter 16)

DISTANCE FUNCTIONS, 118, 124

DOD (DEPARTMENT OF DEFENSE) LOGISTICS, 3 (Chapter 1)

DOMINANCE AND FEASIBILITY, 180 (Section 8.5)

DSA (DEFENSE SUPPLY AGENCY), 266, 269

DUALITY RELATIONS, 124 (Section 7.6)

DYNAMIC PRODUCTION MODELS, 129 (Section 7.8)

DYNAMIC PROGRAMMING, 200, 209, 242, 258, 260, 264, 265, 370, 373, 418, 422

ECONOMETRIC VERSUS GENERALIZED PRODUCTION FUNCTIONS, 102 (Section 7.2)

ECONOMICS, 366 (Bibliography)

EOQ (ECONOMIC ORDER QUANTITY), 248, 268

EPOCH OF SHIFT, 236, 239

ESTIMATORS, 225, 226, 231, 235, 243

EXPERIMENTAL DESIGN, 48

EXPONENTIAL DISTRIBUTION, 230, 231, 237-242, 252, 253, 270, 304, 313, 315, 317, 324-326, 340, 344

EXPONENTIAL SMOOTHING, 228, 229, 270, 272

EXTREMAL METHODS, 351 (Chapter 14)
Strategies for applications, 352 (Section 14.2)

EXTREME VALUE PROBLEMS, 297, 302

FACILITIES DESIGN, 63

FACILITY LAYOUT AND LOCATION, 206 (Chapter 9)

FAILURE RATE, 239

FATHOMING, 135

FIBONACCI SEARCH, 141

FIXED CHARGE PROBLEMS, 372 (Bibliography)

FLSIP (FLEET LOGISTIC SUPPORT IMPROVEMENT PROGRAM), 269

FMSO (FLEET MATERIAL SUPPORT OFFICE), 266

FORECASTING HORIZON, 229

FRONT-END LOGISTICS, 26 (Section 4.3)

GAME THEORY, 353, 359 (Bibliography)

GAMMA DISTRIBUTION, 227, 232, 298, 302, 318

GENERALIZED PRODUCTION FUNCTION, 106 (Section 7.4)

GENERAL PROBABILITY DISTRIBUTION FOR DEMAND, 252

GEOMETRIC DISTRIBUTION, 325

GOOD ALGORITHM, 418, 429

GRAPH THEORY, 373

GREEDY ALGORITHM, 165

HAZARD RATE, 336

HEURISTICS, 54, 137, 142, 143, 152, 174-176, 208, 264, 427, 429

HOLDING COST, 230

ICP (INVENTORY CONTROL POINT), 266, 269

INDIFFERENCE CLASSES, 129

INSPECTION EPOCHS, 241

INTEGER PROGRAMMING, 301, 365 (Bibliography), 416, 418, 419, 421, 425, 430

INVENTORY, 248 (Chapter 11)
- Gap between theory and practice, 248, 249, 272
- Holding costs, 230, 252
- Issue interval, 228
- Lead time, 230, 252, 253, 268, 270
- Procurement costs, 252
- Safety level, 268, 269, 272
- Shortage costs, 230, 252
- Supply system response time, 274

INVENTORY (continued)
Systems, 74-77

INVENTORY DEMAND DISTRIBUTIONS
Binomial, 225
Exponential, 230, 231, 237-242, 252, 253, 270, 304, 313, 315, 317, 324-326, 340, 344
General, 252
Negative binomial, 39, 227, 233, 270
Poisson, 38, 39, 225, 226, 230, 232, 233, 252, 264, 270, 297, 326-328

INVENTORY DEMAND FORECASTING, 223 (Section 10.2)
Bayes approach, 226, 254, 262
Data collection for, 224, 273, 274
Decoupled, 254, 262
Exponential smoothing, 228, 229, 270, 272
General aspects, 40, 223, 224, 265, 270, 273
Low demand, 224 (Section 10.2.1)
Min-max, 254, 262
Not extremely low demand, 227 (Section 10.2.2)

INVENTORY MODELS
Adaptive, 229 (Section 10.3), 303
Approximate, 258, 262, 265, 273
Back ordering, 252
Birthday problem, 296
(Section 12.2)
Categorization, 250
Constraints, 252, 260, 268, 269, 272, 273
Cost discounting, 251
Criteria of optimality, 251, 274, 278
Deterministic, 250
Direct approach, 253, 258
Dynamic, 251
Dynamic programming, 253, 258, 260
Finite horizon, 251
Heuristic, 258, 262, 265, 273
Infinite horizon, 251
Lost sales, 252
Markov chain, 259
Markov process, 236, 253, 258, 259, 264
MIT model, 268
Multi-echelon, 234, 254, 256, 260, 265, 273, 274
Multi-item, 254, 256, 265, 269, 272, 273
Multi-product, 256, 260
Multi-station, 227
Nonstationary, 252
Perishable goods, 250
Production control and, 250
Queuing theory, 259, 262, 265
Renewal theory, 253, 258, 259
Static, 251
Stationary, 252, 258
Stochastic, 250
Transportation modes, 274

INVENTORY POLICIES
Bayes prediction, 234
Centralized control, 40
Continuous review, 251, 256, 257, 265, 268-270, 273
Economic lot size formula, 248
EOQ formula, 248, 268
Joint ordering provision, 260
Periodic review, 251, 256, 257, 259, 260, 265, 272
Play-the-loser, 303 (Section 12.3)
(r,Q), 251
(s,S), 251, 253, 258, 260, 261
Three-and-one rule, 268
Transaction reporting, 251, 274
Trigger point, 251
Types of, 251, 262, 265
Wilson's lot size formula, 248, 268-270

INVENTORY PRACTICE
Army, 269
COSAL, 270
DSA, 269
FLSIP program, 269
General, 248, 265
IBM OS/360 package, 270
Survey of, 265
Theory, 248, 249
VOSL program, 269

INVENTORY RESEARCH
Direct approaches, 253
Other approaches, 253, 258, 262
Time trends in, 255, 265

INVENTORY RESEARCH RESULTS
Analytical, 253, 260
General, 253, 260
Numerical, 253, 260, 272
Policy characterizations, 253, 260

IROS PROGRAM, 27

ISSUE INTERVAL, 228

KUHN-TUCKER CONDITIONS, 402

LAGRANGIAN, 421, 422

LAW OF DIMINISHING RETURNS, 103, 117, 118

L-COM, 42

LEAD TIME FOR INVENTORY RESUPPLY, 230, 252, 253, 268, 270

LEAST SQUARES ESTIMATORS, 226

LEONTIEF MODEL, 102, 106, 113, 114, 366

LIFE DISTRIBUTIONS, 237, 334 (Chapter 13)
Bounds for classes, 342 (Section 13.4)
DFR class, 337
DFRA class, 337
DMRL class, 336
DPL class, 345
IFR class, 336
IFRA class, 336
IMRL class, 337
Models for classes, 340

LIFE DISTRIBUTIONS, (continued) (Section 13.3)
NBU class, 335
NBUE class, 335
NWU class, 336
NWUE class, 336
Preservation of classes, 341, 342
Related classes, 345 (Section 13.6)
Statistical inference, 344 (Section 13.5)

LINEAR PROGRAMMING, 123, 126, 134, 137, 141, 145, 150, 154-156, 160, 164, 167, 170, 171, 175, 184, 186, 188, 190, 191, 198, 354, 357 (Bibliography), 365 (Bibliography), 369, 370, 373, 384, 388, 392, 417, 422, 423

LOGISTICS
Air Force, 25 (Chapter 4)
Army, 11 (Chapter 3)
Decision processes, 52 (Section 5.7), 61 (Section 6.2)
DOD, 3 (Chapter 1)
Effects on weapons research, 21 (Chapter 3)
Front-end, 26 (Section 4.3)
General considerations, 25, 296, 416
Influences on early research, 35, 36
Information for decision-making, 35 (Chapter 5)
Interactive relationships, 25 (Section 4.2)
Managerial functions, 67
Models and (general) decision-making, 30, 52 (Section 5.7)
Navy, 21 (Chapter 3)
Software, 77 (Section 6.3)
Support systems, 29 (Section 4.4)
Systems design, 59 (Chapter 6)
Trade-offs in, 5, 12, 17, 18, 21, 25, 28, 29

LOGISTICS ISSUES AND PROBLEMS
Acquisition, 5, 6, 14, 15, 26
Asset visibility, 16
Automatic data processing, 8
Availability of resources, 3
Centralized versus decentralized, 17
Container ships, 19
Design-to-cost, 6, 15
Direct support systems, 18
Economic impact of defense spending, 8
Effects on weapons research, 21 (Chapter 3)
Elimination of duplication, 9
Facilities management, 7
Initial designs, 23, 26, 28
Integrated logistics support, 22
Intervals between aircraft inspections, 48

LOGISTICS ISSUES AND PROBLEMS (continued) (Section 5.6)
Inventory, 39
Life cycle costing, 6, 16, 27
Macro versus micro in research, 10, 25, 26
Maintainability, 22-24, 26
Maintenance policy, 5, 20, 23, 24, 30, 31, 40 (Section 5.3)
Measures of effectiveness, 39, 40
Overseas bases, 11, 12
Procurement, 6, 26
Production, 6
Provisioning, 28
Range versus depth, 39, 225
Readiness, 3, 4, 13, 14, 21, 29, 30
Reliability, 22-24, 26
Scheduling, 43 (Section 5.4), 65
Ship construction, 129 (Section 7.8)
Size of maintenance facilities, 31, 32
SPS40 air search radar, 23, 24
Stock funding, 17, 18
Supply management, 4, 38 (Section 5.2)
Trade-offs, 12, 16, 17, 21, 28, 30, 31, 42, 44, 52
Transportation, 7, 19
Utilization of manpower, 9
Value engineering, 6

LOGISTICS RESEARCH PROJECT, 224

MAINTENANCE, 40, 306, 334

MAINTENANCE AND MATERIAL MANAGEMENT SYSTEM (3M), 313, 317

MAINTENANCE POLICIES
Age replacement, 334
Block replacement, 334
Comparisons, 337 (Section 13.2)
Replace at failure only, 334

MANAGEMENT SCIENCE, 250

MANPOWER, 306, 331

MARKOV CHAIN, 235, 236, 259, 303, 304, 306, 311
Epoch of shift, 236

MARKOV PROCESS, 235, 236, 264, 327
Stationary, 235

MATURATION PHENOMENON, 50

MAXIMUM LIKELIHOOD ESTIMATORS, 226, 231, 235, 243

MEAN ABSOLUTE DEVIATION (MAD), 270, 272

MEAN SQUARE ERROR EFFICIENCY, 226

MEASURES OF EFFECTIVENESS, 30, 39, 40

MEDICAL, 314

MIT INVENTORY MODEL, 268

MLSF (MOBILE LOGISTICS SUPPORT FORCE), 266

MONOTONE LIKELIHOOD RATIO, 231

MONTE CARLO, 134, 235, 239-241

MULTIPLE REGRESSIONS, 228

NATIONAL INVENTORY CONTROL POINTS, 16

NAVAL PERSONNEL RESEARCH AND DEVELOPMENT LABORATORY (NOW CENTER)
Ship II simulation, 314, 323

NAVAL RESEARCH LOGISTICS QUARTERLY, 250, 351 (Chapter 14)

NAVY OFFICE OF CIVILIAN MANPOWER MANAGEMENT, 355, 356

NEGATIVE BINOMIAL DISTRIBUTION, 39, 227, 233, 270

NEOCLASSICAL PRODUCTION FUNCTIONS, 99 (Section 7.1)

NETWORK THEORY, 175, 208, 209, 363, 373, 383 (Bibliography), 417, 419, 429

NODE SWAPPING, 155

NORM, 52

NORS, 30, 39

OPERATIONS RESEARCH, 250

OPERATOR THEORY OF PARAMETRIC PROGRAMMING, 164, 165, 388 (Chapter 15)
Applications, 396-398, 402, 403, 406
Computational Comparisons, 399, 409 (Section 15.5)
Cost operators, 395, 397-399, 405
For generalized transportation problem, 399 (Section 15.3)
For time transportation problem, 403 (Section 15.4)
For transportation problem, 389 (Section 15.2)
Rim operators, 392, 396, 401
Weight operators, 402

OPTIMIZATION
Strategies for applications, 352 (Section 14.2)

ORNSTEIN-UHLENBECK PROCESS, 327, 330

PARETO OPTIMALITY, 353,

PARETO OPTIMALITY (continued), 398

PERIODIC REVIEW INVENTORY POLICY, 251, 256, 257, 259, 260, 265, 272

PLAY-THE-LOSER RULE, 303 (Section 12.3)

POISSON DISTRIBUTION, 38, 39, 225, 226, 230, 232, 233, 252, 264, 270, 297, 326-328, 341

POLARIS SYSTEM, 224

POMCUS STOCKS, 12

PROBABILITY MODELS, 292 (Chapter 12)

PROBLEMS
- Assignment, 164, 165, 200, 208, 363 (Bibliography), 368, 399, 406, 407, 409
- Decision critical path, 397
- Extreme value, 297, 302
- Facility layout and location, 134, 182, 188, 206 (Chapter 9)
- Fixed charge, 372 (Bibliography)
- Inventory, See inventory entries
- Knapsack, 165
- Layout and location, 206 (Chapter 9), 423
- Maintenance modeling, 334 (Chapter 13)
- Minimax layout, 213
- Minimum spanning tree, 418, 429
- Module placement, 209
- Planar layout, 211
- Quadratic assignment, 208, 368 (Bibliography)
- Scheduling, 134, 373 (Bibliography), 397
- Search and surveillance, 384 (Bibliography)
- Sequencing, 134, 373 (Bibliography)
- Set partitioning, 431
- Shock models, 340
- Statistical, 223 (Chapter 10)
- Transportation, 363 (Bibliography), 372, 383, 388 (Chapter 15), 426
- Traveling salesman, 134, 147-149, 164, 165, 181, 200, 207, 208, 397, 418, 419

PRODUCTION FUNCTIONS, 99 (Chapter 7)
- Axioms, 115-117
- Cobb-Douglas and CES, 106
- Cost and revenue functions, 122 (Section 7.5)
- Disposability of inputs and outputs, 103, 107-110, 114-116, 118, 122, 124
- Duality relations, 124 (Section 7.6)
- Dynamic models, 129 (Section 7.8)
- Econometric forms, 102,

PRODUCTION FUNCTIONS (continued)
103 (Section 7.3)
Econometric versus generalized, 102 (Section 7.2)
Efficient inputs, 101, 117
Exogenous inputs, 107, 110, 112, 115, 117, 130
Generalized neoclassical, 106 (Section 7.4)
Homogeneity, 114, 120, 122, 123
Homotheticity, 119-121, 123, 124, 127, 128
Indirect, 126 (Section 7.7)
Intermediate products, 99, 111-113, 130
Isoquants, 100, 116, 117, 119, 120, 123, 127
Joint, 119
Level sets of, 100, 101
Linear, 100, 102, 103, 106, 109-112, 123, 126
Mathematical forms, 106
Neoclassical, 99 (Section 7.1)
Production paradox, 402
Production possibilities, 99, 100
Semi-homogeneity, 120-122
Statistical estimations of, 102, 105, 106, 128
Substitutions, 100, 102, 105

PRODUCTION PROCESS TAXONOMY, 73

PROGRAMMING
Chance constrained, 369
Combinatorial, 416, 418
Convex, 370
Discrete, 416 (Chapter 16),
Dynamic, 200, 209, 242, 258, 264, 265, 370, 373, 418, 422
Generalized linear, 422, 423
Goal, 355, 356
Integer, 301, 365 (Bibliography), 416, 418, 419, 421, 425, 430
Integer linear, 134, 137, 141, 145, 150, 160, 164, 186, 188, 365 (Bibliography)
Lattice, 253
Linear, 123, 126, 134, 154-156, 160, 164, 167, 170, 171, 175, 190, 191, 354, 357 (Bibliography), 370, 373, 384, 388, 392, 417, 422
Linear fractional, 370
Linear under uncertainty, 369
Mathematical, 370 (Bibliography, 376 (Bibliography)
Mixed integer, 420-424
Mixed linear, 134, 154, 184, 186, 198
Network, 417
Nonlinear, 273, 370, 417
Nonlinear Integer, 301
Parametric, See operator theory
Probabilistic, 369 (Bibliography), 373

PROGRAMMING (continued)
Quadratic, 368 (Bibliography)
Semi-infinite, 370

PROJECTION, 158, 159

QUADRATIC AND QUADRATIC ASSIGNMENT PROBLEMS, 368 (Bibliography)

QUEUING THEORY IN INVENTORY MODELING, 259

QUICKEST DETECTION THEORY, 237

READINESS, 3, 4, 13, 14, 21, 29, 30
Operational, 223, 235 (Section 10.4)

RENEWAL THEORY, 237, 253, 258, 259, 324, 325, 335, 338

REPAIR, 322

REPLACEMENT TIME, 240

RTOK, 47

(r,Q) POLICY, 251

SAMSOM, 42

SCHEDULING, 43 (Section 5.4), 65, 134, 145, 160, 162, 163, 177, 188, 192, 195, 200, 373 (Bibliography), 397, 432

SEMI-POSITIVE VECTOR, 115

SEQUENCING AND SCHEDULING PROBLEMS, 373 (Bibliography)

SHADOW PRICES, 125, 126

SHORTAGE COST, 230, 270

SIM (SELECTIVE ITEM MANAGEMENT), 270

SIMULATION, 42, 134, 235, 239-241, 354, 373, 385
Ship II, 314, 323

(s,S) POLICY, 251, 253, 258, 260, 261

STATISTICAL CONTROL, 229-231, 234

STATISTICAL INFERENCE, 344, 345

STATISTICAL PROBLEMS, 223 (Chapter 10)

STATISTICAL SAMPLING, 13, 51

STOCK LEVEL, 230

SUPPLY, See inventory entries

SURVEILLANCE, 223, 241-243, 384

SURVIVAL FUNCTION, 335, 346

SYSTEMS
Bounds on mean life,

SYSTEMS (continued)
343, 344
Coherent, 340
Design, 59 (Chapter 6)
Interactive, 53, 54
Management control, 53, 54

TECHNICAL COEFFICIENTS, 107, 108

TECHNOLOGICAL PROGRESS, 105, 106

TOLERANCE LIMITS, 233

TRANSPORTATION CORPS, 14

TRANSPORTATION PARADOX, 392

TRANSPORTATION PROBLEM, 363 (Bibliography), 372, 383, 388 (Chapter 15), 426
Assigning uses to sources, 396, 402
Computational comparisons, 399, 409 (Section 15.5)
Generalized, 388, 399
More than one objective function, 398
Optimal growth path, 396
Row-column sum method, 363
Standard, 388-390
Stepping stone method, 363
Stochastic generalized, 402
Time (bottleneck), 388, 403, 405

TRAVELING SALESMAN PROBLEM, 134, 147-149, 164, 165, 181, 200, 207, 208, 397, 418, 419

UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATOR, 235

UPPER CONFIDENCE LIMITS, 233

VOSL (VARIABLE OPERATING AND SAFETY LEVEL) PROGRAM, 269

WEAROUT DETECTION, 223, 227

WEIBULL DISTRIBUTIONS, 239, 240, 242, 300

WILSON'S LOT SIZE FORMULA, 248, 268

WRSK (WAR READINESS SPARES KITS), 28, 29

V. Mathematical Programming: A. Charnes, W. W. Cooper, E. Reed III, Gerald L. Thompson, and Jerrey F. Shapiro.

The book's range may be gauged from this listing of some of the topics discussed: logistics aspects of weapons research, the organization of information for logistics decision making, cost and production functions, applications of branch-and-bound methods in scheduling, inventory theory and practice, classes of life distributions useful in maintenance modeling, network methods in logistics research, computational results for transportation problems, and applications of integer and combinatorial programming in logistics.

Also from The MIT Press

Operations Research for Public Systems
edited by Philip M. Morse
assisted by Laura W. Bacon

"Operations research professionals, noting the success of applications in the industrial and the military sectors, became convinced that many of their methods, such as system modeling, computer simulation, mathematical programming, and the application of the theory of stochastic processes, could be useful in public affairs. At the same time a growing number of managers of public operations and experts in urban and regional planning became aware of operations research and began to be interested in trying out its techniques in solving some of the problems of the public sector, such as those in urban operations, in public health, and in education.

"Operations Research for Public Systems is an attempt to establish communication between the two groups by indicating to the operations research experts the nature of some of the problems in the public sector which might be amenable to present operations research procedures, and by describing to managers and planners of public systems the ways in which operations research [illegible]

Analysis of Public Systems
edited by Alvin W. Drake, Ralph L. Keeney, and Philip M. Morse

This book presents leading recent studies on the application of formal modeling for improved delivery of public services. It very significantly updates and extends the type of material found in Operations Research for Public Systems, which began to organize studies about public systems. Most of the chapters can be read with or without detailed consideration of their technical content. For clarity and compactness, much of the intermediate mathematical detail is referenced to other sources. The editors and authors have striven to make it possible for administrators, who may have limited analytic backgrounds, to use this book to develop their own views on the place of formal analysis in system planning operation.

"Each of the 23 chapters was specially commissioned for this volume. The first six chapters are introductory; the remainder show how quantitative, 'operations research' type models can be employed in such areas as emergency ambulance transportation, blood bank inventory control, criminal justice systems, university planning, and airport development."
—Journal of Economic Literature