INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

A Xerox Education Company

72-29,867

SHAFTEL, Timothy Lee, 1946-
STANDARD MODULES IN PRODUCT DESIGN.

Carnegie-Mellon University, Ph.D., 1972
Business Administration

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

CARNEGIE MELLON UNIVERSITY

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION

Submitted in partial fulfillment of the requirements for the degree of Ph.D. in Industrial Administration (Operations Research)

TITLE STANDARD MODULES IN PRODUCT DESIGN

Presented by Timothy Shaftel

Accepted by David P. Rutenberg 8 Oct 1971
Thesis Supervisor, for the Committee — Date

Approved by the Dean

R M Cyert 10/11/71
Date

PLEASE NOTE:

Some pages may have

indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

THE USE OF STANDARD MODULES IN PRODUCT DESIGN

by

TIM SHAFTEL

Carnegie Mellon University
Graduate School of Industrial Administration

Abstract

This thesis consists of a group of related papers on the use of standard subassemblies or modules as components for several different products. In the multiple modular design problem, parts are to be grouped into several modules; these modules are in turn used to fulfill the requirements of several different applications. If we restrict ourselves to the production of a single module, the problem will be called the modular design problem. In both cases, the solution to the problem will be integer. The word continuous will be used to imply that the integer constraints have been forsaken.

The first chapter of the thesis is a general introduction of the topic. This chapter includes some history of standardization, an outline of the problem and an indication of previous work done in the area.

Chapters two and three develop the basic model, discuss the application, and indicate heuristic procedures for solving portions of the problem. Chapter two was coauthored with David Rutenberg. In this chapter we discuss the general multiple modular design problem. Chapter three develops a model which aids in allocating slight product variations (made possible by modular design) to several markets.

Chapters four and five were coauthored with Gerald Thompson. These chapters are theoretical in nature--they discuss the continuous modular (chapter four) and multiple modular (chapter five) design problems. In chapter four an efficient algorithm is given for attaining the solution to the single module case. Chapter five extends the results of four to the more general multiple module case. In this problem there exist many local optima; a heuristic search procedure is outlined which can be used to locate a "good" local solution.

Chapter six discusses an algorithm which would solve the modular design problem (for the integer solution). In this procedure, the solution to the continuous problem is very helpful. The algorithm is expected to be able to solve moderate size problems--stopping the algorithm before completion should yield good solutions to larger problems. Slight modifications of this algorithm could be used to achieve integer solutions to the multiple modular problem. Given the continuous solution to the multiple modular design problem from chapter five, good solutions to the integer problem could be achieved heuristically.

TABLE OF CONTENTS

		Page
CHAPTER 1:	Introduction	
A.	A History of Standardization	1-1
B.	Single Firm Standardization	1-5
C.	Background Material	1-10
D.	Scope of the Thesis	1-12
	References	1-14
CHAPTER 2:	Product Design: Subassemblies for Multiple Modules (coauthored with D. Rutenberg)	
I.	Introduction	2-2
II.	Problem Statement, Notation and Terminology	2-5
III.	Solution Procedure	2-8
IV.	Heuristic Search	2-11
V.	Sample Problem and Results	2-14
VI.	Conclusions	2-19
	Appendix - An Upper Bound on the Number of Modules	2-21
	References	2-24

Page

CHAPTER 3: Product Modifications for New Multi-National Markets

I. Introduction 3-1

II. The Model 3-6

III. Solution Procedure 3-9

IV. Example Problem 3-12

V. Iterative Solution to Example Problem 3-16

VI. Conclusions and Discussion 3-21

References 3-24

CHAPTER 4: A Simplex-Like Algorithm for the Continuous Modular Design Problem (coauthored with G. L. Thompson)

1. Introduction 4-1

2. Duality in the Modular Design Problem 4-2

3. The Restricted Problem and its Solution 4-6

4. The Modified Problem 4-16

5. The Algorithm 4-20

6. Examples 4-23

7. Conclusions 4-29

References 4-31

Page

CHAPTER 5: The Continuous Multiple Modular Design Problem (coauthored with G. L. Thompson)

1. Introduction 5-1

2. The Primal and Dual MMD Problems 5-2

3. Characterization of Local Optim. 5-8

4. Solution Procedures for Guillotine Partitions 5-14

5. Determining the Number of Modules 5-18

6. Examples 5-19

7. Conclusions 5-22

References 5-24

CHAPTER 6: An Integer Approach to Modular Design

A. An Integer Search Based on the Constraint Set 6-1

An Integer Search for the Exact Optimum 6-2

Reduction of the Search Through Masking and Dominance 6-3

Difficulties with Non-Integer Solution Procedures Applied to Integer Modular Design Problems 6-4

Conclusions 6-5

B. An Integer Search Using the Continuous Solution 6-7

A Lower Bound of Partial Integer Solutions 6-7

Upper and Lower Bounds 6-8

The Algorithm 6-11

Heuristic Schemes for a Good Integer Solution 6-14

Conclusions 6-16

References 6-18

Parts of the Thesis Which Have Been Presented or Published

Chapter 2 (coauthored with D. Rutenberg) has been presented at the 18th International Meeting of TIMS in Washington, D. C., March 1971. It has been accepted for publication by Management Science.

Chapter 3 will be presented at the 12 American Meeting of TIMS in Detroit, September 24 to October 2, 1971.

Chapter 4 (coauthored with G. L. Thompson) will be presented at the 40th National Meeting of ORSA in Anaheim, California, October 27-29, 1971.

Part A of Chapter 6 (under the title of that chapter) has appeared in Operations Research, January-February, 1971.

Acknowledgements

I would like to acknowledge the aid of Doctors Robert Kaplan, David Rutenberg and Gerald Thompson. Their abilities and personalities pervade the pages of this thesis.

CHAPTER 1: Introduction

A. The History of Standardization

The virtues of setting specifications and the advantages of producing items, similar along many dimensions, have been generally recognized since the earliest civilizations. It is not particularly odd that Egyptian architects (ca. 3500 B.C.) recognized that the hewn blocks used in pyramid construction required specific dimensions if cataclysmic destruction, of edifice and career, were to be avoided. Nor is it odd that a group of Babylonian elitists (ca. 4000 B.C.) should recognize the power wielded in the standards of time for a strictly agricultural economy. It is also not strange to discover that standards when set up by diverse and non-interacting social structures should become very much non-standard outside of the local community. In the event of diverse standards it was often necessary for some collusion among societies or for decree by some powerful will in order to create standards amongst such units. Napoleon, for instance, spread the use of the metric system throughout all of continental Europe.

Quantification - languages, weights, and measures - were the first attempts at standardization. They remain, of course, an important backdrop for any discussion of standardization. Governmental attempts at maintenance of standards may be traced to King Shulgi of Babylonia (ca. 2350 B.C.) who had a testing house located in the capital to certify official weights. The Athenians, being international traders, used their own standards of weight as well as those of the Persians and the Phoenicians. The intent behind most standard systems was taxation. In ancient Palestine, for instance, excellent government-owned potteries produced standard vessels for measuring taxes paid in honey and olive oil. Official Arabian (ca. 800 A.D.) cups and bowls bore

the stamp of the governor certifying that they were up to government standards.

The advent of movable type (ca. 1400 A.D.) marked the initiation of a new concept of standardization. That of interchangeability. A small number of basic units could now be used and reused in a variety of combinations. Interchangeability easily found its way into governmental occupations. During the fourteenth century the Arsenal of Venice proclaimed that all bows had to be made so that any arrow could fit them. Further application of this concept led to near assembly line production of galleys at the same arsenal. By 1483, a monk travelling to the Holy Land was to compalin that the galleys were all " ... so much alike ..." [12].

In the U. S., Eli Whitney is credited with being the first to combine the aspects of interchangeability and mechanical production techniques. The acclaim of Sir Joseph Whitworth by the British and of General De Gribeuval and Lablanc by the French (not to mention the recognition of a second American gunsmith, Simeon North) might indicate that Whitney's "invention" was, in 1793, more on the order of an observation. Nevertheless, Whitney may be accredited with "getting the government interested." Contracted to deliver 10,000 muskets to the U. S. War Department, Whitney in 1801, completely assembled ten guns from parts chosen at random by a panel of military men. Impressed by Whitney's legerdemain, the War Department was happy to overlook the fact that only ten muskets were being delivered -- proving, at least, Whitney's ability as a businessman.

Organized attempts at inter-industry standardization in the U. S. started at the Franklin Institute. In 1864, a special committee endorsed a study by William Sellers on screw thread suggesting that the pitch standard should be

placed at 60°. It is ironic that pitch standards were set at 55° in England in 1841 under the guidance of Joseph Whitworth. The result of Sellers' fiasco was to delay standardization between the two countries in this area for eighty-four years and is indicative of the ethnocentric approach to standardization which international societies are attempting to combat. Other standardization controversies during this period surrounded the railroads. Standard guage rails, signalling systems, railway accounts and even standard time were initiated between 1863 and 1900.

The twentieth century saw the advent of multiple efforts by groups of nations - particularly in the area of warfare where coalitions demanded it. The advantages of standardization became obvious on a national and international level. Hundreds of engineering standardizations groups arose between 1900 and 1918, the most notable being the American Institute of Electrical Engineers Standards Committee. In 1918, the AIEE spawned the American Engineering Standards Committee which became the American Standards Association in 1928 --a federation of technical organizations. In general, the ASA does not write or initiate standards; it merely serves as a forum and can approve standards. It is a specific principle of the ASA that standards must be a consensus of those involved and that compliance is strictly voluntary. The ASA is a member of the International Organization of Standards.

Meanwhile governmental agencies were springing up in most countries. England established the National Physical Laboratory in 1900. Congress authorized the formation of the National Bureau of Standards in 1901. Although the NBS works closely with the ASA, its primary functions are not closely related to product designs. Rather, the NBS is charged with setting national standards of weights and measures, determining physical constants,

setting up standard tests and establishing health and safety criteria. In the U. S., therefore, standardization is left to the individual industries under the loose guidance of the ASA.

B. Single Firm Standardization

Within a single firm, product design is closely identified with the marketing function. This identification results from the stress which marketing groups place on new product introduction -- the design of new products to fill the specified demands of various market segments [11]. The approach is to first determine the required characteristics on which to base a product, then to design the product to fit these characteristics. Thus American automobile industries will design "compact cars" to fill an observed demand for small and inexpensive automobiles. The design stipulations in this instance will be to meet the characteristics of the observed (or predicted) demand. Smallness and inexpensiveness would be two criteria while safety and color might be other "marketable" dimensions to be considered.

Figure 1 indicates the process by which a product is ultimately designed. Traditionally, finance and production functions in design have been limited to the consideration of feasibility. Finance, for instance, might discourage a new product because of high initial costs during a tight money market. Production might eliminate a product due to lack of required machinery and tooling. The changes in product design initiated by production and finance result as a by-product of the conflict between subgoals of the various groups involved. March and Simon [13] indicate four major processes by which an organization reacts to conflict: (1) problem solving, (2) persuasion, (3) bargaining, and (4) "politics." The final outcome will depend largely upon the differences between the various subgoals as well as the relative strengths and weaknesses of the marketing and production groups and their personnel [20].

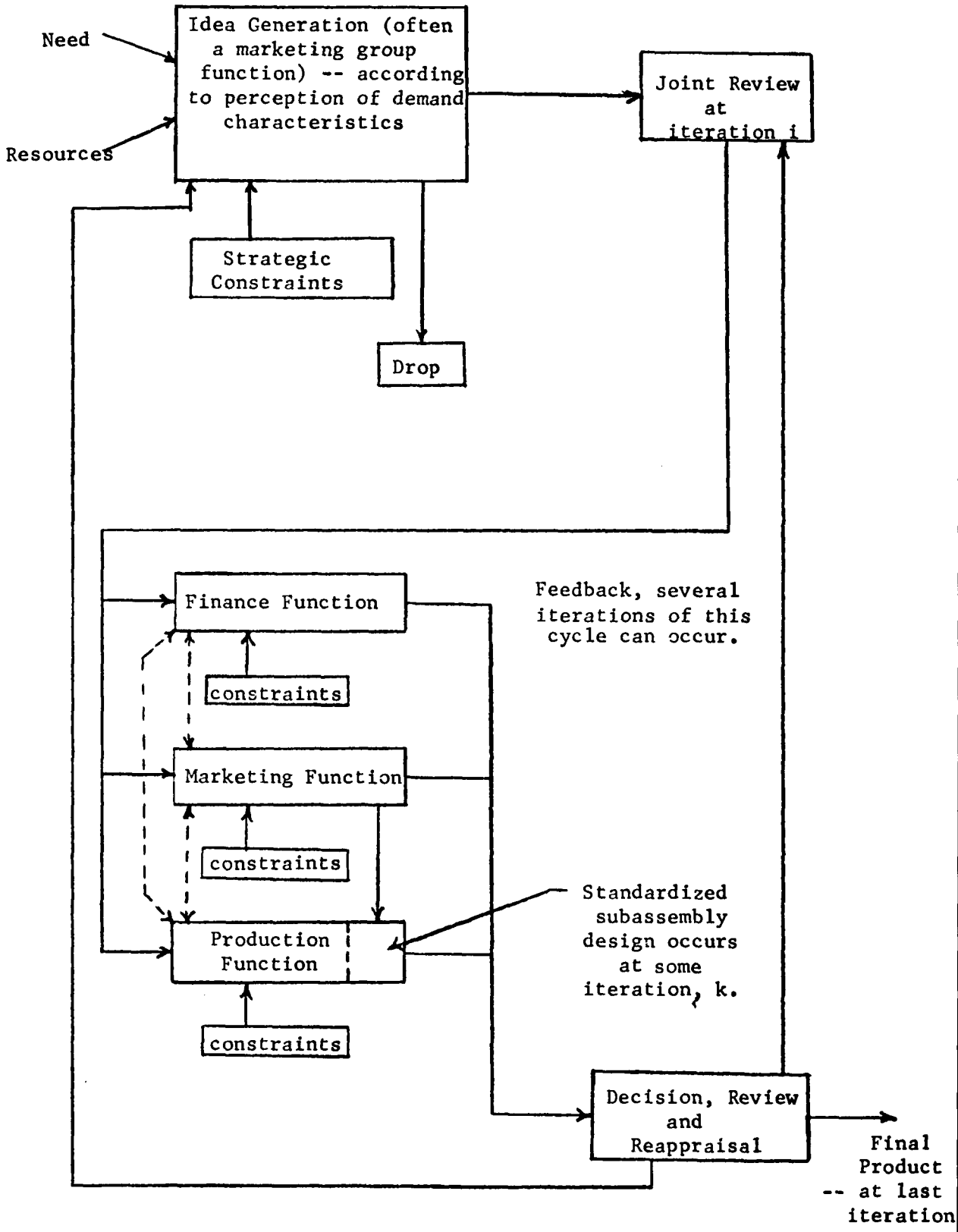


Figure I: Product Design

It has been rare to find the production function in design to be considered as more than merely fulfilling the criteria of the marketing function. Only recently has the use of standard designs (modules) been considered in attempting to ffect production and inventory costs. To be sure, Tinker Toys and Erector Sets have long used a very few types of standard parts which fit together in a multitude of ways. In these instances, however, the design stipulations were a market decision (i.e., to fill a demand for "put-together" toys); they were not a result of production planning for minimum cost. Westinghouse Electric used few types of standardized circuit boards in its second generation computer, Prodac 500, but only because of decisions made outside the production group on bases other than cost. Not until the past few years has product design been attempted as a production function in standardizing subassemblies. Indeed, only five years ago, Martin Starr was led to write an article with the illuminating title, "Modular Design -- A New Concept" [22].

Modular design implies more than a consideration of the required demand dimensions. It means more than a standardization because standard designs are in some sense "sellable." Modular design is the choice of the number of standard designs to be produced, the choice of how many of each type of part to be included in each of the various designs, and the choice of how many of each standard design should be used in each application. All three choices must be made simultaneously with an eye to overall cost minimization. It is used to develop a product in such a way as to minimize the joint costs of production, inventory, consumer disutility, repair, and maintenance. To illustrate this process, consider the following example:

"An electronics corporation has four markets -- one in Canada, a second in Mexico, and two in the U.S. (plant production and overhaul market). Each of the four markets has a known demand for two types of TV's, three types of radios, and two types of electronic calculators that the corporation produces. Each of these applications has a specific requirement for three basic parts: inductors, resistors, and capacitors -- each of which has an associated cost. The company can produce up to a specified maximum number of different circuit boards or modules [9],[10]. Each module contains some combination of the basic parts. (The number of parts is limited by reliability and ease of replacement considerations [6].) There is a fixed cost involved in setting up the production of any one module and in designing test equipment. Replacement modules must be in stock in any market where the modules are used; there is a fixed cost, therefore, of carrying any one module in inventory in a market. Finally, there are certain costs related to the number of parts on a module and the number of modules used in an application. These costs include transportation costs (a unit cost which is a function of the market and the weight of the module) and handling costs (a function of the number of modules used for the several applications). The problem is to minimize costs by determining the number of modules to build, the number of each part to place in the various modules, and which modules to use in each application in each market."[19]

Figure II is a list of the advantages and disadvantages of producing standard subassemblies.

Advantages	Example
1. Fewer types of standard designs must be inventoried in any one market.	Programming packages often contain several subroutines which can be combined to form a multitude of programs. Storage space for a few subroutines is much smaller than storing several complete programs.
2. Fewer types of standard parts provide tighter quality control and standard testing procedures.	Producing bigger quantities of the same circuit board means that standard testing procedures can be created. The costs of developing these procedures are spread over a greater number of boards.
3. Repair is simplified by ease of replacement.	Repairmen need carry only a few types of standard modules to quickly repair appliances.
4. Cannibalization (the use of parts from one application to repair a second) is simplified [7].	Standard modules removed from an unrepairable aircraft can be used to repair a radar van.
5. Additions to a product line can be simplified.	New generation computers can be manufactured from standard subassemblies already designed and tested.
6. Adaption to market segments is simplified.	Use of options can alter an automobile's market segment.

Disadvantages	Example
1. More parts than required are used.	Excess nuts and bolts in an assembly kit are thrown away.
2. Excess parts increase cost of transportation and handling.	Nuts and bolts which are not necessary are packaged and shipped at some cost.
3. Interconnection of modules may be difficult.	"Backboard wiring" of modules in a computer is time consuming and costly in terms of reliability.

Figure II: Advantages and Disadvantages of Standard Subassemblies

C. Background Material

The first mathematical work on modular design was done by David Evans in 1963 [4]. Evans formulated the problem in terms of producing only one style of standard module and attempted to minimize the cost of excess parts in a deterministic approach to the problem.

Evans' formulation was:

$$\text{Min} \sum_{i \in I} c_i x_i \sum_{j \in J} d_j y_j$$

$$\text{s.t.} \quad x_i y_j \geq r_{ij} \qquad \text{For all i and j}$$

$$x_i, y_j, c_i, d_j \geq 0$$

where $I = \{1,2,\ldots,m\}$

$J = \{1,2,\ldots,n\}$

c_i = cost of part i

d_j = demand for application j

x_i = the number of part i on the module (decision variable)

y_j = the number of modules in application j (decision variable)

Evans developed an algorithm to solve the problem using a special non-linear programming approach. His approach was facilitated by the similarity of the constraints and the objective function. Evans, however, ignored any integer requirements. In 1965 A. Charnes and M. Kirby [2] developed a new solution procedure for Evans' formulation, using generalized inverses and convex programming. Applying the transformations $c_i x_i \equiv e^{u_i}$ and $y_j d_j \equiv e^{v_j}$, they formed the following equivalent problem:

$$\text{Min} \sum_{i \in I} \sum_{j \in J} e^{u_i + v_j}$$

$$\text{s.t. } u_i + v_i \geq c_i d_j \ln r_{ij} \qquad \text{For all i and}$$

The result was a linear constraint set and a convex objective function. By using an algorithm tailored to this problem, a solution is reached. The integrality conditions were again overlooked. In 1970, U. Passey wrote a third article [15] on modular design using Evans' formulation. Passy solved the problem using geometric programming.

In 1969, David Rutenberg formulated and solved the complementary problem to modularity [18]. His formulation focuses on the production of many modules to be used one at a time in a variety of applications. Rutenberg called this problem commonality. Combining the concepts of commonality and modularity leads quite naturally to the production of more than one module to be used, several at a time, in a number of applications.

D. Scope of the Thesis

This thesis is a collection of papers dealing with the topic of modular design. The relationship between these papers can be seen from the diagram on the following page.

Chapter II was written jointly with David Rutenberg. It is a formulation of the overall problem (multiple modules) outlined in Part B. This paper also presents a heuristic solution to the formulation given. Chapter III is a formulation of the problems of allocating variants of a specific product to several markets. A solution procedure to this model is also given. Chapters IV and V were jointly authored by Gerald Thompson. Chapter IV goes back to the original (Evans) formulation of the modular design problem (using a single module). In this paper an algorithm for solving the continuous problem is given. This algorithm solves large modular design problems quickly and efficiently. Chapter V uses the results of Chapter IV to determine the continuous solution to the multiple module design problem. It provides a procedure for identifying local optimum solutions and gives a heuristic approach to finding a good local optimum. Chapter VI discusses algorithms and heuristics for finding the integer solution to modular design problems once the continuous solution is found. Adaptations of these procedures could be used to solve the multiple modular design problem for integer solutions.

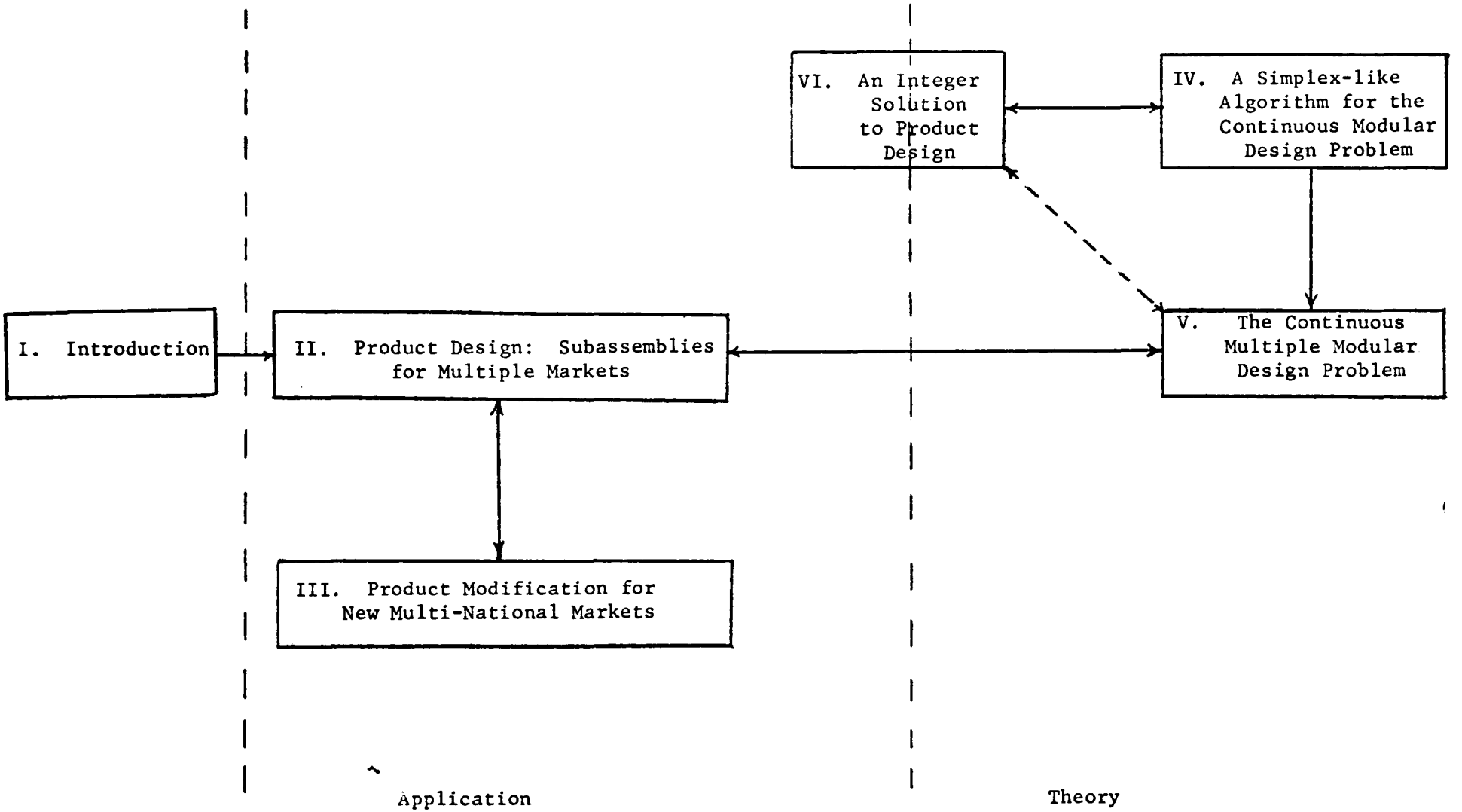


Figure III: Theory and Application

References

[1] Brady, Robert, Organization, Automation, and Society, University of California Press, 1963.

[2] Charnes, A. and M. Kirby, "Modular Design, Generalized Inverses and Convex Programming," Operations Research, Vol. 13, 1965, [pp. 836-847].

[3] Coles, Jessie, Standards and Labels for Consumers' Goods. The Ronald Press Company, 1949.

[4] Evans, David H., "Modular Design -- Special Case in Non-Linear Programming," Operations Research, Vol. 11, 1963, [pp. 637-647].

[5] Gaillard, John, Industrial Standardization, Its Principles and Application, The H. W. Wilson Company, 1934.

[6] Greenman, L. R., "Packaging Requirements for Optimized Malfunction Isolation by Systematic Substitution," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-5, No. 3, May 1969, [pp.499-514].

[7] Hirsh, M., Meimer, M. and Boll, C., "Cannibalization in Multi-Component Systems and the Theory of Reliability," Naval Research Logistics Quarterly, Vol. 15, September 1968, [pp. 331-359].

[8] Industrial Standardization, Studies in Business Policy, No. 85, National Industrial Conferences Board, June, 1957.

[9] Kantz, William H., "Cellular Logic-in-Memory Arrays," IEEE Trans-actions on Computers, Vol. C-18, No. 7, July 1969, [pp. 719-727].

[10] Koenig, W. A., "Universal Component Packaging System," IEEE Transactions on Parts, Materials and Packaging, Vol. PMP-1, No. 1, June 1965, [pp.230-236].

[11] Kotler, Philip, Marketing Management, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1967.

[12] Lane, Frederic Chopin, Venetian Ships and Shipbuilders of the Renaissance, Johns Hopkins Press, 1934, [p. 17].

[13] March, J. and Simon, H., Organizations, Wiley, New York, 1958.

[14] Melnitsky, Benjamin, Profiting from Industrial Standardization, Conover-Mast Publications, Inc., 1953.

[15] _____, Management of Industrial Inventory, Conover-Mast Publications, Inc., 1951.

[16] Passy, U., "Modular Design: An Application of Structured Geometric Programming," Operations Research Vol. 18, May 1970, [pp. 441-453].

[17] Reck, Dickson (ed.), National Standards in a Modern Economy, Harper and Brothers, Publishers, 1956.

[18] Rutenberg, David P., "Design Commonality to Reduce Multi-Item Inventory: Optimal Depth of a Product Line," Operations Research, Vol. 19, 1971, [pp. 491-509].

[19] Rutenberg, David P. and Shaftel, Tim, "Product Design: Subassemblies for Multiple Markets," Management Science [to appear].

[20] Simon, Herbert A., "On the Concept of Organizational Goals," Administrative Science Quarterly, June 1964, [pp. 1-22].

[21] Simplification, Standardization, Specification; Case Studies on Variety Reduction, Vol. 1, The European Productivity Agency of the Organization for European Economic Co-operation, 1958.

[22] Starr, Martin K., "Modular Production -- A New Concept," Harvard Business Review, November-December 1965, [pp. 131-142].

[23] Struglia, Erasmus J., Standards and Specifications, Sale Research Co., 1965.

Management Sciences Research Report No. 207

PRODUCT DESIGN: SUBASSEMBLIES FOR MULTIPLE MARKETS

June 1970

by

David P. Rutenberg

and

Timothy L. Shaftel

Management Sciences Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

PRODUCT DESIGN: SUBASSEMBLIES FOR MULTIPLE MARKETS

David P. Rutenberg

and

Timothy L. Shaftel

Abstract

A corporation may profit by installing more parts than its products require. Each product or application requires a specified number of parts such as inductors, resistors and capacitors. For assembly and maintenance reasons, parts are grouped into subassembly modules. If there are economies of scale in manufacturing subassembly modules, it may be worthwhile to standardize on a few module designs that will be used in all products or applications even though a few more parts than needed are installed. The cost of giving away extra parts by using few standard modules must be balanced against the fixed costs of producing more types of standard modules. In the multi-market problem there is a fixed cost with the production of a module, and then another fixed cost when the module is used in a market. The paper presents a solution procedure and the time results of computer runs for the problem of finding the best standard modules for a multi-product, multi-market corporation.

PART I

Introduction

Product design is an art which reconciles aesthetics, technical performance, and cost. Products are rarely designed from scratch; they are usually modifications and syntheses of preceding generations of designs. The design team, therefore, has at least a ball park idea of the demands of the principal market segments. When a design team calls for management science aid, it envisions one of three options:

1. Information Retrieval and Display Systems

   Market research reports, lost sales reports, customer complaints, and competitive surveys are available within most corporations but are rarely digested for a design group. One information retrieval system is in use by Sulzer Bros., the Swiss designer of large marine diesel engines. This system encompasses the manufacturing capability of each of its licensees around the world, the non-availability of certain standard items (such as square tubing), and maintenance reports on each engine.

2. Automated Design Programs

   Certain products are designed to the specifications of each customer; the design procedure is standard and, therefore, automated. Westinghouse Electric, for example, uses computer programs in designing small and medium-sized electric motors to issue instructions for each set of windings. More complex design problems can also be handled [7].

3. Optimizing Design Standards [this paper]

   This paper shall consider the use of standard designs--how many different types should be produced and the specifications of each type.

This latter problem is particularly challenging in a global corporation where multimarket considerations cannot be ignored. The H. J. Heinz Company sells more than 57 varieties. Indeed, for each product, a different recipe is tailored to each nation. In contrast, the Campbell Soup Company moved international by selling exactly the same product in each nation of the world [13]. Between these two extremes lies an optimum product design. If designs are tailored to the unique specifications of each market segment in each nation, and with an eye to their

customs classification and hence rate of import duty, the amount of consumer dissonance or disutility will be minimized.

By having few products, on the other hand, a company achieves economies of scale from mass production, from reductions in inventories, and from purchasing discounts. Other gains would include the reliability of having more than one plant producing an item and the flexibility of being able to divert product flows between nations with little paper work and few erroneous shipments. A reconciliation between tailored designs and standard designs lies in modularity and commonality of design; many subassemblies would be standard throughout the world, others standard to groups of nations, and yet others tailored to each segment of each different nation. Let us now be more precise.

Modularity is the production of a single module to be used, several at a time, in a number of applications. An example of modular design would be the production of a single parts package to be used, in various numbers, to satisfy parts requirements of several different types of model kits. The modular design problem has been approached for continuous variables by Evans [3] and by Charnes and Kirby [2]. For the case of integer variables, a small problem has been solved by Shaftel [11] using an implicit search procedure. Commonality, on the other hand, is the production of several modules to be used singly for a variety of applications; fewer modules would be produced than there are applications. An example of commonality would be the assortment of different lengths of steel bolts carried as standard items. Sadowski [10] and Rutenberg [9] have worked in this area.

Combined, modularity and commonality would be the production of more than one module to be used, several at a time, in a number of applications. An example of modularity and commonality combined would be the production of a few types of circuit boards, various numbers of which would be used to fulfill the needs of each of a number of different types of electronic apparatus.

The combined problem will be the focus of this paper. Furthermore, we shall handle the multi-market problem. Let us now present the sample problem that will be solved later in the paper:

An electronics corporation has four markets--cne in Canada, a second in Mexico, and two in the U.S. (plant production, and the field repair market). Each of the four markets has a known demand for two types of TV's, three types of radios, and two types of electronic calculators that the corporation produces. Each of these applications has a specific requirement for three basic parts: inductors, resistors, and capacitors--each of which has an associated cost. The company can produce up to a specified maximum number of different circuit boards or modules [5],[6]. Each module contains some combination of the basic parts. (The number of parts is limited by reliability and ease of replacement considerations [4].) Because of required inventory considerations in the four markets, a fixed cost is incurred by a module being used in any market. Finally, there are certain costs related to the number of each part in the modules used in the various markets. These costs include transportation costs (a function of the market and the weight of the module) and handling costs (a function of the number of modules used for the several applications). The problem is to minimize costs by determining the number of modules to build, the number of each part to place in the various modules, and which modules to use in each application in each market.

PART II

Problem Statement, Notation and Terminology

In each market there is a known demand for a number of applications. Each application j requires at least r_{ij} of part i. Because there are economies of scale in manufacture (and joint economies over the n_m markets), the parts will be aggregated into n_k modules, which are then used as subassemblies in each application. In addition to the cost of parts, and the economies of scale in module manufacture, there may be a disutility in using module k in application j for market h, to be represented by the handling and transportation costs.

The notation and terminology used in this paper will be:

n_m = number of markets, $h = 1,2...n_m$

n_p = number of parts, $i = 1,2,...n_p$

n_a = number of applications, $j = 1,2,...,n_a$

n_k = maximum possible number of modules, $k = 1,2,...,n_k$

d_j^h = demand for application j in market h

r_{ij} = requirement of part i in application j

x_i^k = the number of part i in module k (decision variable)

y_j^{kh} = the number of module k used in application j for market h (decision variable)

$z^{kh} = \begin{cases} 0 \text{ if module } k \text{ is not used for market } h \\ 1 \text{ if module } k \text{ is used for market } h \end{cases}$ (decision variable)

$z^k = \begin{cases} 0 \text{ if module } k \text{ is not used} \\ 1 \text{ if module } k \text{ is used} \end{cases}$ (decision variable)

g^h = the maximum number of modules allowed in market h (decision variable in heuristic search)

c^i = cost of part i

f^h = fixed cost incurred when a module is used in market h

f = fixed cost incurred when a module is produced

$u_j^{kh}(x_i^k, y_j^{kh})$ = disutility cost of using module k in application j for market h (a function of x_i^k and y_j^{kh})

w_i = weight of part i

s_j^h = cost per unit weight for shipping product j to market h

b_j = cost of fitting a module in application j

Using this notation, the objective function which we wish to minimize is:

$$\theta = \min \sum_k \left[\sum_i c_i\, x_i^k \sum_j \sum_h d_j^h\, y_j^{kh}\right] \quad \text{cost of parts}$$

$$+ \sum_k \sum_h f^h\, z^{kh} \quad \text{fixed cost of using module } k \text{ in market } h$$

$$+ \sum_k f\, z^k \quad \text{cost of setting up module } k$$

$$+ u_j^k(x_i^k, y_j^{kh})\Big\} \quad \text{disutility cost as a function of the number of modules and the number of parts on a module (transportation and handling costs).}$$

where,

$$u_j^k(x_i^k, y_j^{kh}) = \sum_k \sum_j \sum_h b_j\, d_j^h\, y_j^{kh} \quad \text{handling costs}$$

$$+ \sum_k \sum_j \sum_h \sum_i d_j^h\, s_j^h\, y_j^{kh}\, w_i\, x_i^k \quad \text{transportation costs}$$

After combining terms, the problem with constraints becomes:

$$\theta = \min \sum_k \sum_j \sum_h d_j^h\, y_j^{kh} \left\{\sum_i (c_i\, x_i^k + s_j^h\, w_i\, x_i^k) + b_j\right\}$$

$$+ \sum_k \sum_h f^h\, z^{kh} + \sum_k f\, z^k$$

$$\text{s.t.} \quad \sum_k x_i^k\, y_j^{kh} \geq r_{ij} \qquad \forall_{i,j,h} \qquad (1)$$

$$\sum_j y_j^{kh} - z^{kh} M \leq 0 \qquad \forall_{k,h} \qquad (2)$$

$$\sum_h z^{kh} - z^k_M \leq 0 \qquad \forall_k \qquad (3)$$

$$z^{kh} = 0 \text{ or } 1 \qquad \forall_{k,h} \qquad (4)$$

$$z^k = 0 \text{ or } 1 \qquad \forall_k \qquad (5)$$

All terms are positive and x_i^k, y_j^{kh} are integer.

Constraint (1) assures that the number of parts in any application will meet or exceed its requirements. Constraint (2) (along with (4) and (5)) assures that a module is inventoried in a market where it is used while constraint (3) assures that module's production.

PART III

Solution Procedure

The problem consists of three sets of variables, x, y, Z. The computational procedure for each will be discussed, in the order x (components per module), Z (selection of modules for each market application) and finally y (number of each module in each market application). The procedure may be viewed as a two-stage dynamic program. In the first stage x is set; then, in the second stage Z and y are optimized.

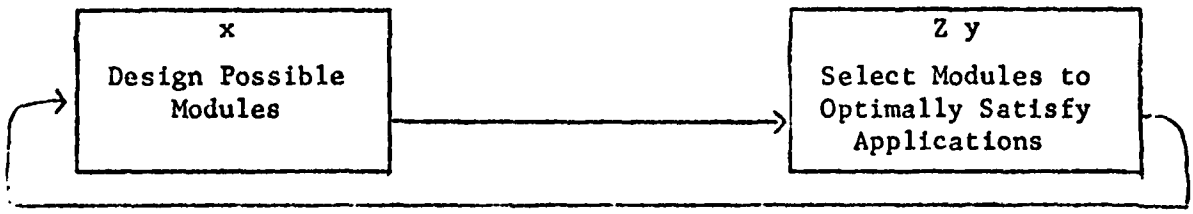


Figure 1: Stage 1, Stage 2 of Computational Procedure

Note that by setting the module design $x_i^k = \bar{x}_i^k$, and then the selection of modules for each market application $Z^{kh} = \bar{Z}^{kh}$ (which in turn set $Z^k = \bar{Z}^k$ by constraint 3), we are left with a set of $n_m \cdot n_a$ small subproblems in y:

$$\theta_{jh} = \min \sum_k q_j^{kh} y_j^{kh}$$

$$\text{s.t.} \quad \sum_k \bar{x}_i^k y_j^{kh} \geq r_{ij} \quad \forall_i \qquad \forall_{j,h}$$

$$y_j^{kh} \in I \geq 0 \quad \forall_k$$

where, $q_j^{kh} = d_j^h \left\{ \sum_i (c_i \bar{x}_j^k + s_j^h w_i \bar{x}_i^k) + b_j \right\}$

Each subproblem has n_k unknowns, n_p constraints, and

$$\theta(\bar{x}_i^k, \bar{Z}^{kh}, \bar{Z}^k) = \sum_j \sum_h \theta_{jh} + \sum_k \sum_h f^h Z^{kh} + \sum_k f Z^k$$

for any fixed values of x_i^k and Z^{kh} (and, therefore, Z^k).

The procedure can be followed in detail by using Figure 2.

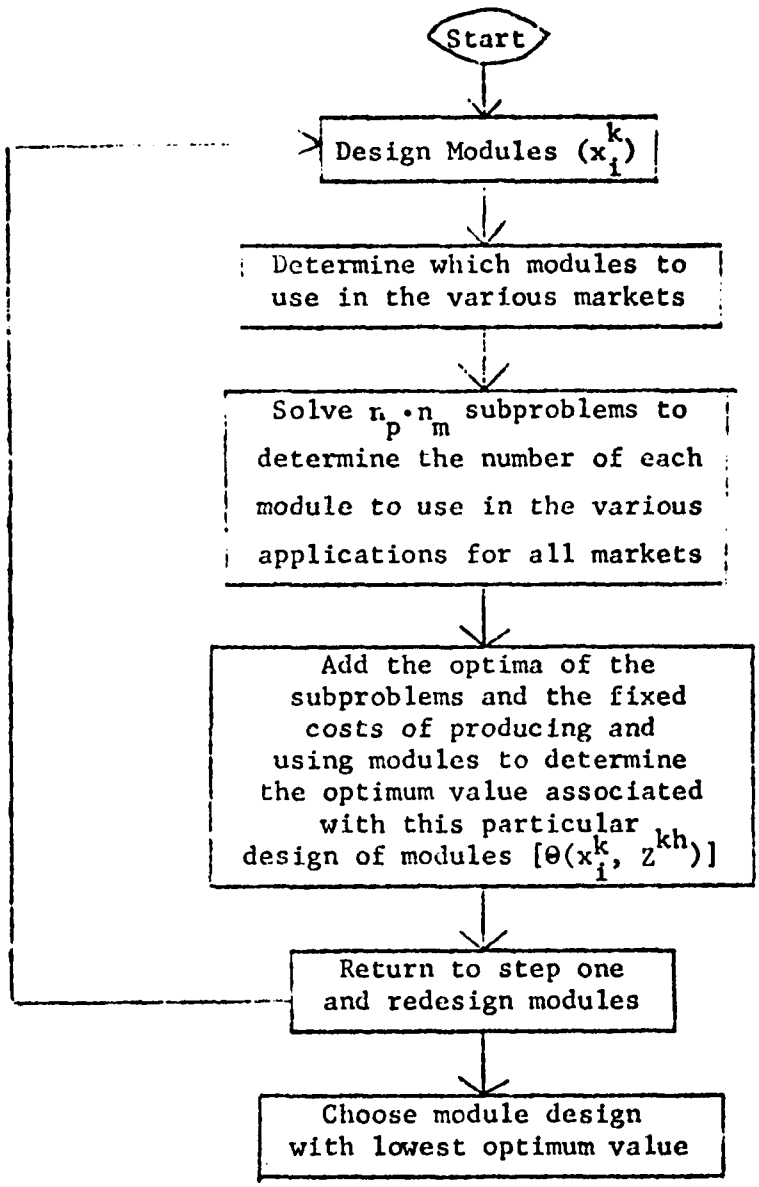


Figure 2: Flow Chart of Solution Procedure

No module would be optimal if it contained more of any part than the greatest number of that part needed by any application; hence, an upper bound on x_i^k is $\max_j r_{ij}$ for all i. An upper bound on the maximum number of modules, n_k, which need to be designed can also be derived (see Appendix). Subject to these two bounds, an exhaustive search of the entire feasible region would find the exact optimum. However, the number of enumerations required would be

$$\left(\prod_{i=1}^{n} \max_j r_{ij}\right)^{n_k} \cdot 2^{n_k \cdot n_m}$$

where $n_k \geq n_a$ in all cases (as can be seen from the upper bound on n_k developed in the Appendix). This is too many enumerations for any but the smallest of problems; a heuristic search procedure will be presented which will yield results for much larger problems.

PART IV

Heuristic Search

The heuristic search will follow the format of the exact procedure just cutlined. We shall present heuristics for use in designing the module; selecting the modules to be used; and determining the number of each module to be used in each application in each market.

1) Designing the Modules (x_i^k). In this stage we specify the number of each component to be fitted into each module, and also set g^h, the maximum number of modules allowed in a market h. Associated with each vector (whose elements are x_i^k and g^h) is the optimized objective function of the second stage problem. We are therefore dealing with a surface having dimensions of the vector x and g.

For this surface a number of minimization procedures are available. Given the not perfect convexity of the surface (see Figure 3), we selected a particularly robust procedure developed by Box [1] for the Imperial Chemical Industries Limited. The procedure is initialized by generating (from insight or randomly) points in the space we are searching. The number of points generated should be one greater than the dimensions of the space. In our case, each point will correspond to the design of n_k modules and the statement that no more than g^h can be used in market h. The value of the objective function associated with each point has to be calculated as outlined in Sections 2) and 3) below.

At a typical iteration the worst point (parent) of those available at this time is selected. This parent will be discarded at the end of the iteration. The centroid of the remaining points is then calculated and a new point is generated at some distance, $\alpha(> 1)$, on the opposite side of the centroid from the parent (the new point must then be rounded to depict whole components). If the new point has the highest cost of those in the set, a second point is created half-way between the parent and the centroid. If a point remains worst for four iterations, we modify the Box procedure to drop the point and randomly generate a new point

as its replacement. This modification eliminates the possibility of the algorithms getting "stuck" on a ridge of the objective function.

This portion of the heuristic could be improved as soon as there is more insight into the shape of the objective function. If the non-convexity can safely be ignored, we could use the points to find the minimum of an implicitly fitted quadratic function, or use difference equations to improve on the Box procedure. Furthermore, a search that explores only integer points would be beneficial.

2) Selecting the modules for each market (Z^{kh}). The Box procedure used in stage one stipulated g^h the maximum number of modules to be used in each market. In the exhaustive ideal case all $\binom{n_k}{g^h}$ combinations of g^h modules out of the n_k available would be considered. The second heuristic procedure, however, calculates the centroid of the n_k possible modules and then eliminates those modules closest to the centroid, until only g^h remain for market h. If the module remains, $Z^{kh} = 1$; if the module is eliminated, $Z^{kh} = 0$. The order of elimination is the same for all markets.

The rationale behind eliminating modules closest to the centroid is that more diversified modules will be available for use. A possible but costly improvement would be to solve the third stage problem as though all n_k modules were available. The g^h modules most often used in each market could then be used to solve the third stage problem again. The advantage from going through the third stage of computation more than once must be considered before the usefulness of this improvement can be determined.

3) Optimizing the Number of Modules used (y_j^{kh}). The modules have been designed and a subset chosen for each market. We must now decide how many of each module to use in each application. We can formulate this decision as integer because only whole modules can be used. In our heuristic procedure we relax

the integer constraint. All values in the solution vector are initially truncated (rounded down) so that this integer solution is no longer feasible. The cheapest module is identified; the number of those cheap modules is then raised by one. If the solution remains infeasible, the next lowest cost module is inceased by one. This procedure continues until the solution becomes feasible.

The obvious improvement in this portion of the heuristic would be to solve the integer programming problem using one of the available integer techniques (Branch and Bound, Gomory Cuts, etc.). Alternatively, the non-integer solution procedure might be improved by usage of the module which yields the largest gain towards feasibility per unit cost.

PART V

Sample Problem and Results

BASE CASE

The numerical example will apply to the verbal problem which introduced the paper. The input data were:

Number of markets, $n_m = 4$ Canada, Mexico, U.S. plant, U.S. maintenance market

Number of kinds of parts, $n_p = 3$ inductors, resistors, capacitors

Number of applications, $n_a = 7$ 2 types of TV's, 3 types of radios, 2 types of electronic calculators

Maximum allowable number of modules, $n_k = 8$

The cost of parts, c_i =

-Part- 1	2	3
.017	.021	.025

The weight of parts, w_i =

-Part- 1	2	3
.0032	.0041	.0010

Fixed cost of module production, f = 3948.00

Fixed cost of using a module in a market, f^h =

-Market- 1	2	3	4
28492	17271	958	9728

Handling charge per module, b_j =

1	2	3	4	5	6	7
.025	.015	.032	.007	.002	.009	.012

Demand for each application in the four markets, d_j^h =

Application \ Market	1	2	3	4
1	118	600	200	840
2	49	850	230	960
3	20	500	590	720
4	41	320	432	240
5	0	100	300	500
6	10	100	0	300
7	100	100	0	500

The requirements by each application for the 3 parts, r_{ij} =

Application \ Part	1	2	3
1	7	8	4
2	9	2	8
3	500	820	940
4	602	735	421
5	100	100	5
6	50	500	10
7	3	5	80

Shipping cost per unit weight to various markets. For each application s_j^h =

Application \ Market	1	2	3	4
1	2.5	3.10	4.07	2.04
2	2.75	3.35	3.90	1.57
3	2.26	3.05	3.34	1.21
4	2.01	1.42	1.38	1.99
5	2.0	3.0	4.0	1.0
6	2.0	3.0	4.0	1.0
7	2.0	3.0	4.0	1.0

We assumed that no module could contain more than 15 of any part due to manufacturing and reliability considerations. More complex module design constraints could be accommodated.

SOLUTION TO THE BASE CASE

The problem was solved for the minimum cost configuration. The optimal solution found involved the production of three modules:.

The number of each part used on the modules, x_i^k =

Module \ Part	1	2	3
1	12	14	15
2	14	2	13
3	4	10	1

Modules 2 and 3 were to be used only in the U.S. plant (market 3). The other markets used one module primarily because of the high carrying cost, f^h, in these markets. The number of modules used for this solution were:

Number of modules 1, 2, and 3 used in each application, y_j^{kh}

Market \ Application	1	2	3	4	5	6	7
1	(1,0,0)	(1,0,0)	(63,0,0)	(53,0,0)	*	(36,0,0)	(6,0,0)
2	(1,0,0)	(1,0,0)	(63,0,0)	(53,0,0)	(9,0,0)	(36,0,0)	(6,0,0)
3	(0,1,1)	(0,1,1)	(58,6,0)	(6,21,61)	(0,5,10)	*	*
4	(1,0,0)	(1,0,0)	(63,0,0)	(53,0,0)	(9,0,0)	(36,0,0)	(6,0,0)

*Application not sold in this market

Note that the numbers of each module used were identical in the markets which required only one module. The similarity results from the near equality of the ratio of costs from market to market. This fact implies that parametric programming might be very advantageous for this problem, and appropriate heuristics could be developed.

The cost for the optimum solutions found was 2.907×10^5. The ratio of this solution to the lower bound (the fixed cost incurred by any solution--developed in the appendix) was 1.22.

STEP LENGTH AND NUMBER OF ITERATIONS

The step length in the Box routine, α, was set at 1.15 for this problem after evaluating several alternatives between 1.0 and 1.3. The number of iterations was set at 1,000; 200 iterations, however, appeared almost as effective. These values, of course, must be set with consideration of the size of the search space. The number of parts, the maximum possible number of modules, and the upper bound on the number of parts on any one module appear to be the critical factors which influence the decision.

SHAPE OF THE OBJECTIVE FUNCTION

Figure 3 (a, b, and c) indicates the shape of the objective function around its optimum value while holding all variables constant except for the number of parts 1, 2, and 3 on module one (i.e., x_1^1, x_2^1, x_3^1 are allowed to vary). The near convexity of these curves is encouraging when considering modifications to the Box search algorithm.

EXECUTION TIME STUDY

A group of 13 sample problems was run for 200 iterations using various values for n_k, n_m, n_a, and n_p in order to test execution time on a Univac 1108. (The biggest job tested used $n_k = 16$, $n_m = 4$, $n_a = 7$, $n_p = 5$.) Using the minimum squared error fit through these points yields the predictive equation:

$$\text{time/200 iterations} = 0.088 \cdot n_k^{0.95} \cdot n_m^{0.99} \cdot n_a^{0.91} \cdot n_p^{1.77} \text{ seconds } (R^2=0.98).$$

Note that with this heuristic algorithm we expect that the computer running time will remain reasonable for commercially sized problems.

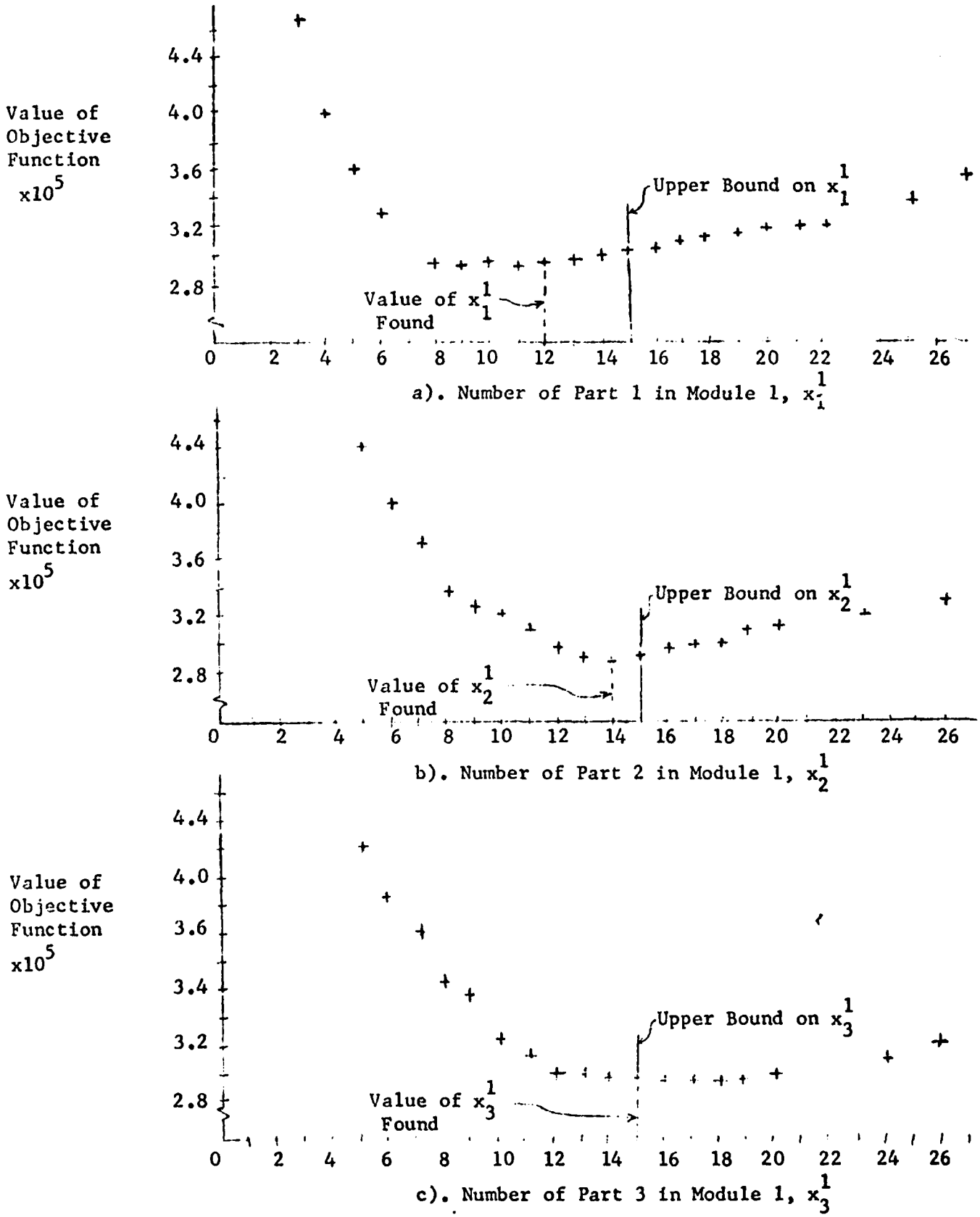


Figure 3 a,b,c. The Shape of the Objective Function Around the Optimum Value Found

PART VI

Conclusions

In this paper we have merged the problem of modularity with that of commonality, and have expanded the combined problem to multiple markets, as faced by a multinational company. We have also developed and tested a solution procedure (albeit a heuristic) which appears fast enough to solve commercial problems. In addition, methods have been outlined which will make the solution procedure more precise, though these would involve longer computer runs. The paper is intended to aid the design of products with complex subassemblies though it appears to have broader applicability than just electronic apparatus. While aesthetic aspects of product design must be affirmed, we have merely designed subassembly modules that achieve technical feasibility at low cost. Let us detail other bounds on the paper.

The paper presents a static analysis. Market demands, however, grow as products pass through their life cycles. Modules appropriate for the demands of one moment in time should also be designed with an eye to their place in the line of modules after demands have increased. This problem will be of more concern to some corporations than to others.

The paper presents a deterministic analysis. One never knows real demands with certainty. Furthermore, one would like the flexibility to incorporate technological innovations as they occur. The forseeable can conceptually be modelled by stochastic programming with recourse [14]. The unforseeable can be handled by having this solution procedure available in order to quickly evaluate possible tactics.

In a multinational corporation the modules may be manufactured in different locations and then imported into their nations of usage. Import duties often depend on the customs classification, which depends on the module design. Constraints for each classification can be imposed on the choice of modules;

the handling cost can then include the relevant import duty plus freight. This approach still presupposes a decision as to where modules are to be manufactured. For each set of locations, the model can be re-run.

There is no claim that this model solves the entire problem of product design. Rather we hope that this model will be useful in the design of complex products. We believe that it will lead to greater flexibility and adaptability as many cases are run to explore possible states of the world.

Appendix - An Upper Bound on the Number of Modules

Theorem 1--An upper bound on the objective function, θ, can be given by the expression:

$$\theta' = \sum_i \sum_j \sum_h d_j^h (c_i r_{ij} + s_j^h w_i r_{ij}) + \sum_j \sum_h b_j d_j^h + N_a (f + \sum f^h)$$

This is the objective function which occurs when a separate module is produced for each application.

Proof: Let $x_i^k = r_{ij}$ if $k = j$ $\forall$ $i,j,k=1,\ldots,n_a$

and $y_j^{kh} = 1$ if $k = j$

$= 0$ otherwise

$\forall$ $j,h,k=1,\ldots,n_a$

Constraint 1 is satisfied since $x_i^k y_j^{kh} = r_{ij}$ for $j = k$. Constraints 2 and 4 imply that $z^{kh} = 1$ for all k and h; constraints 2, 3, and 5 imply that $z^k = 1$ for all k. Substituting these values into the objective function yields θ'.

This upper bound corresponds to the case where each module is made to fit the exact requirements of a particular application. The number of modules produced is then equal to the number of applications and all of these modules are used in each market.

Theorem 2--A lower bound on the objective function value can be given by the expression:

$$\theta'' = \sum_i \sum_j \sum_h c_i d_j^h r_{ij} + \sum_i \sum_j \sum_h d_j^h s_j^h w_i r_{ij} + \sum_h \sum_j b_j d_j^h + f + \sum_h f^h$$

which corresponds to the case where one module is produced which exactly fulfills the requirements of all applications.

Proof: The minimum value of $\sum x_i^k y_j^{kh}$ is r_{ij}. Substituting this value into the objective function yields the first two terms of θ''. The last three terms

come from the fact that one is the smallest number of modules which can be used in any application for each market. Note that this bound also represents a fixed cost incurred by any feasible solution.

Theorem 3--The maximum number of modules which must be considered is

$$n_k = \left\langle \left\{ n_a (f + \sum_h f^h) - \sum_{\substack{h \\ h \neq \ell}} f^h \right\} / (f + f^\ell) \right\rangle$$

where $f^\ell = \min_h f^h$ and $\langle V \rangle$ denotes the smallest integer greater than V.

Proof: From theorem 2 any solution which produces one module must have an objective function of, at least, θ''. If we have n_e extra modules produced (after the initial one) the objective function becomes θ''' at minimum, where

$$\theta''' = \theta'' + n_e f + n_e f^\ell$$

and $n_k = n_e + 1$ is the number of modules produced. θ''' is indicative of the fact that n_e extra modules were produced and that they were, at least, used in the market of minimum cost per module, market ℓ. Since extra modules will be produced until θ''' exceeds a known, feasible objective function value, we have, using Theorem 1:

Whenever $\theta''' \geq \theta'$, extra modules are no longer profitable, which will be true iff

$$\sum_i \sum_j \sum_h c_i d_j^h r_{ij} + \sum_i \sum_j \sum_h d_j^h s_j^h w_i r_{ij} + \sum_h \sum_j b_j d_j^h + \sum_h f^h + (n_k - 1) f + (n_k - 1) f^\ell$$

$$\geq \sum_i \sum_j \sum_h c_i d_j^h r_{ij} + \sum_i \sum_j \sum_h d_j^h s_j^h w_i r_{ij} + \sum_h \sum_j b_j d_j^h + n_a (f + \sum f^h)$$

$$\Leftrightarrow \quad n_k f + n_k f^{\ell} + \sum_{h \neq \ell} f^h \geq n_a (f + \sum_h f^h)$$

$$\text{or} \quad n_k \geq \frac{n_a (f + \sum_h f^h) - \sum_{h \neq \ell} f^h}{f + f^{\ell}}$$

$$\therefore \quad n_k = \left\langle \frac{n_a (f + \sum_h f^h) - \sum_{h \neq \ell} f^h}{f + f^{\ell}} \right\rangle$$

REFERENCES

1. Box, M. J., "A New Method of Constrained Optimization and a Comparison with Other Methods," The Computer Journal, Vol. 8, pp. 42-52 (1965).

2. Charnes, A., and M. Kirby, "Modular Design, Generalized Inverses and Convex Programming," Operations Research, Vol. 13, pp. 836-847 (1965).

3. Evans, David H., "Modular Design--Special Case in Nonlinear Programming," Operations Research, Vol. 11, pp. 637-647 (1963).

4. Greenman, L. R., "Packaging Requirements for Optimized Malfunction Isolation by Systematic Substitution," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-5, No. 3, pp. 499-514 (May 1969).

5. Kantz, William H., "Cellular Logic-in-Memory Arrays," IEEE Transactions on Computers, Vol. C-18, No. 7, pp. 719-727 (July 1969).

6. Koenig, W. A., "Universal Component Packaging System," IEEE Transactions on Parts, Materials and Packaging, Vol. PMP-1, No. 1, pp. 230-236 (June 1965).

7. Mitroff, Ian I., "Simulating Engineering Design: A Case Study of the Interface Between the Technology and Social-Psychology of Design," IRE Transactions on Engineering Management (December 1968).

8. Rosenbock, H. H., "An Automatic Method for Finding the Greatest or Least Value of a Function," The Computer Journal, Vol. 3, pp. 175-184 (1960).

9. Rutenberg, David P., "Design Commonality to Reduce Multi-Item Inventory: Optimal Depth of a Product Line," Operations Research, to appear.

10. Sadowski, Wieslaw, "A Few Remarks on the Assortment Problem," Management Science, Vol. 6, pp. 13-24 (1959).

11. Shaftel, Tim, "An Integer Approach to Modular Design," Operations Research, to appear.

12. Starr, Martin K., "Modular Production--A New Concept," Harvard Business Review, pp. 131-142 (Nov.-Dec. 1965).

13. Sommers, Montrose and Jerome Kernan, "Why Products Flourish Here, Fizzle There," Columbia Journal of World Business, pp 89-97 (March-April 1967).

14. Walkup, D. W. and Roger Wets, "Stochastic Programs with Recourse," SIAM Journal of Applied Mathematics, Vol. 15, No. 5, pp. 1299-1314 (September 1967).

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Graduate School of Industrial Administration
Carnegie-Mellon University

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
Not applicable

3. REPORT TITLE

Product Design: Subassemblies for Multiple Markets

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Management Science Research Report June 1970

5. AUTHOR(S) (First name, middle initial, last name)
David P. Rutenberg
Timothy L. Shaftel

6. REPORT DATE
June 1970

7a. TOTAL NO. OF PAGES
24

7b. NO. OF REFS
14

8a. CONTRACT OR GRANT NO.
NONR 760(24)

b. PROJECT NO.
NR 047-048

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Management Science Research Report No. 207

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
Not applicable

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES
Not applicable

12. SPONSORING MILITARY ACTIVITY
Logistics and Mathematical Statistics Br.
Office of Naval Research
Washington, D. C. 20360

13. ABSTRACT

A corporation may profit by installing more parts than its products require. Each product or application requires a specified number of parts such as inductors, resistors and capacitors. For assembly and maintenance reasons, parts are grouped into subassembly modules. If there are economies of scale in manufacturing subassembly modules, it may be worthwhile to standardize on a few module designs that will be used in all products or applications even though a few more parts than needed are installed. The cost of giving away extra parts by using few standard modules must be balanced against the fixed costs of producing more types of standard modules. In the multi-market problem there is a fixed cost with the production of a module, and then another fixed cost when the module is used in a market. The paper presents a solution procedure and the time results of computer runs for the problem of finding the best standard modules for a multi-product, multi-market corporation.

PRODUCT MODIFICATIONS FOR NEW MULTI-NATIONAL MARKETS

by

Tim Shaftel

May 1971
The University of Arizona
College of Business and Public Administration

ABSTRACT

An international corporation has several product-market strategies available to it. One such strategy is that of moving existing products or minor modifications of existing products into new foreign markets. The chief difficulties of such a move are to determine which variant should be used in a market and the associated optimum transfer or sales prices. The costs include fixed costs of producing new product variations, shipping, inventory, and tariff costs, as well as advertising and distribution costs. Legal, cultural and environmental considerations must also be examined. This paper presents a model which helps develop an optimum marketing strategy. A solution procedure is presented which can solve large problems and which can be modified to allow other factors (such as cross elasticity of demand) to be considered.

I. Introduction

A. Product-Market Strategy

New product development and the identification of new markets is a vital operation for any corporation [25]. Identification of new markets involves searching in new geographic areas, including foreign countries [37], as well as extensive review of present markets and their segmentation [13], [33]. Analytical techniques for segmenting markets along various social, demographic and behavioral characteristics have met with moderate success [6], [8], [14]. Segmentation along obvious geographic and ethnic lines remains, of course, a dominant marketing approach--an approach which has been substantiated in many cases [3], [24]. New product development involves both major and minor modifications to existing products [28],[30], as well as the development of totally new products [5]. Product-market strategies are, of course, not isolated from each other. It will be convenient in this discussion to consider the strategy matrix of figure I.

	Modification of Existing Products	Development of New Products
Segmentation of Present Markets	A	B
Addition of New Markets	C	D

Figure I: Product-Market Strategies

Strategy A is indicative of a decision to modify an existing product in order to take advantage of a different or unheeded [22] market segment.

American automobile manufacturers, for instance, will offer various options to take advantage of perceived market segmentation, at the same time developing a new design every five years because of changing consumer tastes. The risk to the corporation can be minimized with this strategy since minor changes involve few large scale adaptations to production procedures. This strategy is further enhanced by the availability of vast backlogs of data for similar products in similar markets.

Development of new products to fulfill the needs of a market segment already familiar to the corporation (strategy B) can be triggered by either the corporation or the market. Many corporations actively seek new product demand in familiar markets. The Bell Telephone Company, for instance, has an extensive research program to seek out new products in the communications industry. Other corporations only react to stimulus given by the market or some segment of it. Many companies keep close watch on customers' comments in order to identify the possibilities of new products. Risks involved with using this strategy are greater than those involved with strategy A for we are dealing with an untested product. Greater testing costs will be incurred as well as initial costs of production setup and inventory. Sunk costs, if the product is given up late in its development, could be large. Some advantage might be made of promotional "spill-over" [7] from an old product to a new one. An IBM computer carries a great deal of prestige simply because of its name. It should be recognized that this spill-over effect could have either a positive or negative effect.

Strategy C represents the expansion into new markets. The rationale behind such a strategy is that if a product sells in one market, it has a good chance of selling in another. Care must be taken with this assumption [34]; minor modifications of the product may be greatly beneficial. Risks caused by this strategy are reduced by the fact that

product data in at least one market is known. New production is not necessary so that risks can be greatly reduced by market testing without incurring production set up costs. Advertising spill-over from one market to the next must also be considered although positive and negative effects are, once again, both possible.

The final possible strategy (D) is the most risk prone. Introduction of a new product into a previously unknown market involves extensive testing costs and other costs associated with initial production and setup. This strategy, like C, would require an active search on the part of the corporation into new markets and the initiative to produce new products. Substantial argument may be given against this approach as a short term strategy--the risks are high and the possibility of gains questionable, particularly in light of using strategy A,B, or C. A strategy of this sort, however, can be profitable in the long term. A corporation might move from a single product in a single market (i.e. no strategy of product development at all) to strategy A (market segmentation and product modification). As the possibility of product modification increases, new markets (strategy C) may also be considered. New products (strategy B) can be considered as market segments and shifting market demands are recognized. Once new markets become familiar and new products are developed, going to the strategy of new products in new markets would be reduced to using products, known in some markets, in markets which are themselves familiar because of their penetration by other products.

B. Multi-National Markets

The expansion into multi-national markets accentuates the difficulties which must be faced by a corporation [10],[15],[20]. Some degree of market segmentation and product modification will probably be necessary; differences in cultural backgrounds and mores must be considered. While the corporation focuses on its own national markets, the available market information is greatly increased by the managers background. He is working in an area where his intuition and personal experiences can be brought to bear directly on his marketing strategy. In many cases, he can be considered as part of the market he is serving. In foreign markets that same manager must take considerable pains to be sure that full consideration is given to items usually taken for granted in the home country. Take, for example, the following note from an English newspaper [18, p. 135]:

> Sir -- My grocer is neither fool, knave nor profiteer and has served me faithfully for two decades. I now find that a garish establishment opposite is selling my particular tea at 5d less per quarter and I am tempted to "rat" across the road. Taking the long view, however, I think I will change my tea and stick to my grocer. There is something vaguely sinister and un-English about this cut-price racket.

The extent to which demographic and socioeconomic differences occur between various countries offers a strong case for the concepts of market segmentation among countries. Legal requirements and differing excise rates increase the need for product modifications. The shift from 86.8 to 86 proof, for example, reduced Seagrams' U.S. tax by 1.3 million dollars in 1957 [32]. In moving into new international markets the lowest risk strategy is C, using existing product modifications from the home market and attempting to use information from home market segments.

Consideration of some segments of Belgian markets, for instance, might give insight into attempted advancement into French markets. As the foreign market becomes better known to the corporation more or new products can be introduced into that market under much less risk. The use of foreign nationals as advisors and managers must not, of course, be precluded. It must be emphasized that the ability of a corporation to enter new foreign markets can be greatly benefitted by a product which can be easily modified to more closely fit demand characteristics of the various markets. Whether the product prospers or fails may be entirely determined by the custom classification into which it falls or by a coloring acceptable in one country but unacceptable in another.

The next section will give a formulation of a model to determine the use of existing or proposed product modifications in various new markets (strategy C). Although the initial concern will be on existing product modification designs, the model can be used to lend insight into the direction of new modifications.

II. The Model

The model proposed here is based on strategy C, that of moving minor modifications of existing products (called variants) into new markets.

There exist n different variants of a particular product and m different markets. For each market, and each variant, a demand curve or price strategy [29, third strategy] is known. Costs include those which are a function of the quantity of each variant sold in a market such as shipping costs, taxes, duties and advertising expenditures; (in this model these costs are assumed to be linear). Variable (linear) production costs are incurred which are a function of the total amount of each variant produced. Along with variable costs, fixed costs may be associated with production, inventory, duties and advertising. Constraints include legal restrictions such as price and quantity levels as well as environmental restrictions. The problem is to discover which variants, if any, should be marketed in each of the several markets and the associated price.

Notation:

n = number of different variants, $i=1,\ldots,n$

m = number of different markets, $j=1,\ldots,m$

p_{ij} = price of variant i in market j

q_{ij} = quantity of variant i sold in market j ; a function of p_{ij}

c_{ij} = linear costs of shipping, tariffs and inventory for variant i in market j

b_i = production cost per unit of variant i

M_j = marketing budget in market j (includes distribution and advertising)

U_j = fixed inventory cost if any variant is sold in market j

T_j = fixed tariff incurred after a certain quantity of a product enters market j

K_j = upper limit on imports into market j (or point where T_j is incurred)

B_i = fixed cost of producing variant i

$\underline{P}_{ij}$, $\overline{P}_{ij}$ = lower and upper bounds on price for variant i in market j

z_{ij} = 1 if variant i is used in market j
0 if variant i is not used in market j

The objective function of the mathematical formulation is to maximize profits (revenue from those variants sold less the variable and fixed costs of producing and selling those variants):

$$\underset{p_{ij},q_{ij},z_{ij}}{\text{Max}}\ \pi_o = \sum_i \sum_j (p_{ij}q_{ij} - c_{ij}q_{ij} - b_i q_{ij} - M_j - U_j)\, z_{ij} - \sum_i B_i \lambda_i - \sum_j T_j \delta_j$$

where q_{ij} is a function of p_{ij} .

The constraints are designed to perform several tasks. Constraints (1) and (5) (see below) indicate that we, at most, allow one variant per market. Constraints (2) and (3) indicate that as soon as the quantity of a certain variant in market j exceeds a value, K_j, then δ_j becomes one thus incurring tariff T_j. T_j may be infinitely large making K_j an upper limit on the quantity in that market (knapsack constraints). Constraints (4), (5) and (6) assure that the fixed cost, B_i, of producing variant i is incurred as soon as that variant is sold in any market. Constraint (7) is indicative of price regulations. And constraint (8) allows for the elimination of certain variants in certain markets by legal, cultural or environmental considerations. The constraint set is:

(1) $\sum_i z_{ij} \leq 1$ For all j

(2) $\sum_i q_{ij} - K_j - \delta_j L \leq 0$ For all i, j ; L = a large number

(3) $\delta_j = 0$ or 1 For all j

(4) $\sum_j z_{ij} - \lambda_i L \leq 0$ For all i

(5) $z_{ij} = 0$ or 1 For all i, j

(6) $\lambda_i = 0$ or 1 For all i

(7) $\underline{P}_{ij} \leq p_{ij} \leq \overline{P}_{ij}$ For all i, j

(8) $z_{ij} = 0$ For some i and j

III. Solution Procedure

Inspection of the mathematical formulation of this problem shows that, once we have determined the variant we wish to allocate to a market our pricing policy is strictly determined by that market. Indeed, it is possible to determine the optimum return from any market strategy before making the decision of which variant to use in a market. For any market, ℓ, and variant, k, we have:

$$\text{Max} \quad \pi_{k\ell} = (p_{k\ell} - c_{k\ell} - b_k)\, q_{k\ell} - \delta_\ell T_\ell = \pi^*_{k\ell}$$
$$p_{k\ell}, q_{k\ell}$$

where $q_{k\ell}$ is a function of $p_{k\ell}$

$$\text{s.t.} \quad q_{k\ell} - K_\ell - \delta_\ell L \leq 0 \qquad (1a)$$

(A)

$$\underline{P}_{k\ell} \leq p_{k\ell} \leq \overline{P}_{k\ell} \qquad (2a)$$

$$\delta_\ell = 0 \text{ or } 1 \qquad (3a)$$

Constraint 1a is determined from constraint 2 of the main problem since the use of variant ℓ in market k implies that no other variant will be used (from constraint 1). Constraints 2a and 3a come directly from constraints 3 and 7 of the main problem. In the objective function U_ℓ, M_ℓ, and B_k become sunk costs and need not be considered in determining the pricing policy for the subproblem.

Before proceeding, it is necessary to assume the shape of the demand curve for the various markets. Typically assumed demand curves are differentiable and monotonically decreasing functions of price [16]. Once the demand function is known, we can find the optimum price (assuming constraints 1a and 2a are satisfied with δ_ℓ equal to zero), via differentiation. In this paper, the form of the demand function is assumed as:

$$q = G \exp[-\alpha(p + \gamma)]$$

where G, α, and γ are parameters characteristic of the different markets and variants. (The i,j subscripts are left out for simplification.) For this demand function, the unconstrained solution to problem A is:

$$p_{\ell k}^{opt} = [1 + (c_{\ell k} + b_k) \cdot \alpha_{\ell k}]/\alpha_{\ell k}$$

Often, of course, some constraints will be violated. In this case we would have to look at the return at $\underline{P}_{k\ell}$, $\bar{P}_{k\ell}$ and $\bar{\bar{P}}_{k\ell}$-- which is the price associated with K_ℓ, the point at which tax, T_ℓ , is incurred (See Figure II). Several permutations of the positions of the upper and lower bounds on price may be

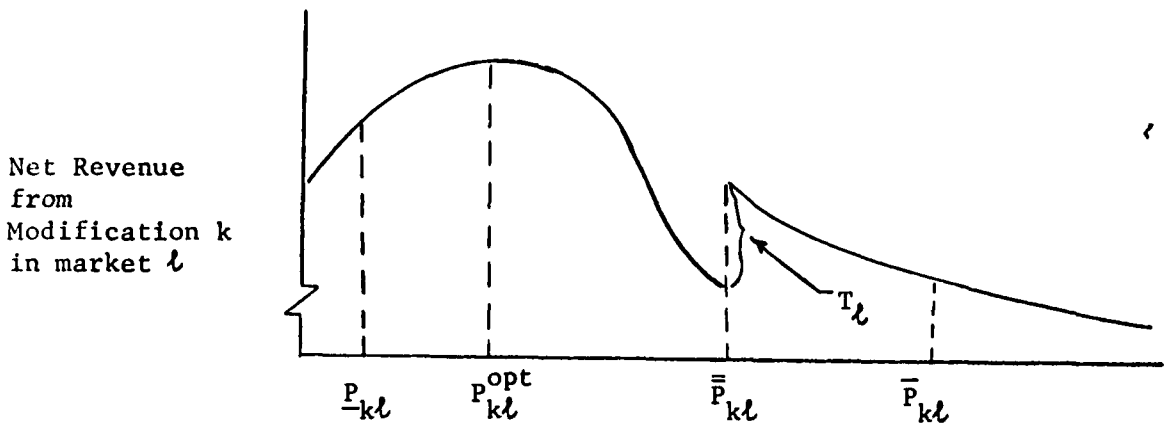


Figure II

made in Figure II. At most we would wish to calculate the revenue at four different prices.

Let $r_{ij} = \pi^*_{ij} - U_j - M_j$, the original problem may now be placed in the following format:

$$\underset{z_{ij}}{\text{Max}} \quad \sum_i \sum_j r_{ij} z_{ij} - \sum_i B_i \lambda_i$$

$$\text{s.t.} \quad \sum_i z_{ij} = 1 \qquad \forall_j \quad (1b)$$

$$\sum_j z_{ij} \leq \lambda_i L \qquad \forall_i \quad (2b) \qquad \text{(B)}$$

$$z_{ij} = 0 \text{ or } 1 \qquad \forall_{i,j} \quad (3b)$$

Constraint 1b is tight since we may eliminate from consideration any market which does not have a profitable product modification.

The final problem is now a simple plant location problem. Several good solution procedures exist for this problem [11], [35]. Efroymson and Ray indicate computer results of ten minutes for 50 markets and 200 variants, a size which would more than likely exceed the size of any practical problem.

IV. Example Problem

Air conditioning is presently an industry with a $5 billion annual installed value in the United States. The 1969 production of unitary air conditioning for commercial and residential use was 1,600,000 units in the United States, 400,000 units in Japan and successively lesser amounts in Italy, Germany and France [1], [2], and [4]. With growth rates of 15 to 20 percent for this industry in developed, foreign countries, it is not surprising that American corporations have attempted to sell in these markets. Borg Warner in Germany; Chrysler's Airtemp in Spain, South Africa and Australia; Copeland Refrigeration Incorporated in Belgium; and General Electric in South Africa are all examples of U.S. corporations moving into international markets.

As an example problem, we will assume that a corporation wants to move into four international markets: Belgium, France, Sweden and South Africa. In order to do this, ten different variants of presently manufactured air conditioners are designed. (Four of these variants are already being produced, therefore incurring no production set up cost.) Certain designs are not acceptable in some countries because of differences in power supplies or because of legal safety requirements. Still other designs are eliminated for certain countries (by a panel of experts) because of cultural considerations. In the matrix which follows a zero in cell (i,j) implies that the use of variant i in market j has been ruled out a priori.

z_{ij}

	Belgium 1	France 2	Sweden 3	South Africa 4
1*	0	0	-	-
2*	0	0	-	-
3*	0	0	0	-
4*	-	-	-	0
5	-	-	-	-
6	-	0	-	-
7	0	0	-	0
8	0	0	-	0
9	0	-	-	-
10	-	-	-	-

* indicates variant exists

\- indicates z_{ij} is free to assume a value of either 0 or 1

The linear costs of variant i in market j are:

c_{ij} =

i \ j	1	2	3	4
1	-	-	60	50
2	-	-	40	40
3	-	-	-	30
4	20	35	45	-
5	30	20	35	40
6	30	-	60	35
7	-	-	30	-
8	-	-	100	-
9	-	25	50	30
10	40	30	40	35

Fixed Cost of Production, B_i

0
0
0
0
1400
1500
1000
2500
4000
1300

	1	2	3	4
Market budget, M_j =	1500	2500	1250	2000
Fixed Inventory, U_j =	2500	2000	1250	2000
Fixed Import Tariff, T_j =	2000	1000	2000	2000
Upper Limit on Imports, K_j =	150	100	1000	500

The upper and lower bounds on the price of variant i in market j

$\underline{P}_{ij}, \overline{P}_{ij}$ =

	1	2	3	4
1	-	-	85,450	40,110
2	-	-	75,150	75,150
3	-	-	-	30,140
4	60,140	50,150	60,100	-
5	50,150	75,150	50, 85	70,200
6	75,200	-	50,275	70,150
7	-	-	75,200	-
8	-	-	60,175	-
9	-	60,130	40,120	60,180
10	60,120	80,110	90,150	70,220

Parameters of the demand curve: g_{ij}, α_{ij}, γ_{ij} =

1	-	-	500, .01, 0.0	600, .015, .005
2	-	-	400, .02, 0.01	600, .014, .004
3	-	-	-	500, .020, .003
4	500, .025, .01	500, .02, .002	350, .025, 0.015	-
5	600, .02, .01	600, .016, .003	500, .025, 0.01	400, .022, 0.01
6	400, .022, .008	-	100, .01, 0.0	350, .018, 0.01
7	-	-	300, .015, 0.0	-
8	-	-	1000, .02, 0.005	-
9	-	750, .015, .004	750, .015, 0.005	750, .02, .002
10	750, .018, .005	750, .02, .004	250, .02, 0.01	600, .025, .00

V. Iterative Solution to Example Problem

The solution to phase I, the calculation of revenue for variant i in market j were:

r_{ij} =

	1	2	3	4
1	-	-	7600	2900
2	-	-	800	5000
3	-	-	-	1050
4	480	50	-820	-
5	2050	4698	580	-1227
6	-545	-	-500	-167
7	-	-	2166	-
8	-	-	0	-
9	-	7166	5260	3550
10	3444	2574	-450	-320

The associated price and quantity for variant i in market j are:

p_{ij}, q_{ij} =

	1	2	3	4
1	-	-	160,101	110,115
2	-	-	90, 66	111,126
3	-	-	-	80,101
4	60,112	85, 91	85, 42	-
5	80,121	112,100	75, 77	85, 61
6	75, 76	-	160, 20	91, 69
7	-	-	97, 70	-
8	-	-	150, 50	-
9	-	92,190	90,194	80,151
10	95,134	101,100	90, 41	75, 92

Solving problem B, for this example, turns out to be quite simple. We are attempting to pick one row entry (variant) in each column (market) so that the sum of the revenues less the sum of the fixed production costs for the variants used is a maximum. We may eliminate initially any row such as row 8 with all negative or zero revenue entries. On problems of this size branch and bound algorithms [e.g., see17, p. 565] may readily be applied. The solution to this problem is:

Market	1	2	3	4
Variant	10	5	1	2
Price	96	112	160	110
Quantity to import	134	100	101	115

Total Profit = $18,042

It is now possible for the designer to inspect these results in order to determine whether or not a new variant should be considered. Using variants 5 and 10 for instance, might lead the designer to produce a new variant (number 11) to be used in both France and Belgium. Assume that variant 11 has the following characteristics:

Cost c_{11j} =

1	2	3	4
35	25	40	40

parameters of demands: G_{11j}, α_{11j}, γ_{11j} =

1	2	3	4
700, .019, .01	680, .017, .003	400, .025, .01	500, .022, .0

	1	2	3	4
Upper and Lower Bound on Price: $\underline{P}_{11j}$, $\overline{P}_{11j}$	55,130	80,135	75,100	70,210

Fixed cost of variant 11 = 1100

The results for this variant are:

	1	2	3	4
revenue: r_{11j} =	2947	4276	-340	-545

	1	2	3	4
Price, quantity: p_{11j}, q_{11j} =	88,132	113,100	80, 54	85, 76

Using this variant in both France and Belgium yields a better optimum. The solution, therefore, will be:

Market	1	2	3	4
Variant	11	11	1	2
Price	88	113	160	110
Quantity to Import	132	100	101	115

Total Profit = $18,723

To complete the example, let us assume that the corporation does not wish to price variant 11 at such a widely different price in Belgium ($88) and France ($113). Note that the quantity to be imported into France is the maximum amount (100) such that the excess quantity tariff is not incurred. By forcing the price in France down, it will become optimal to import more than 100 units and absorb this tax (See Figure III).

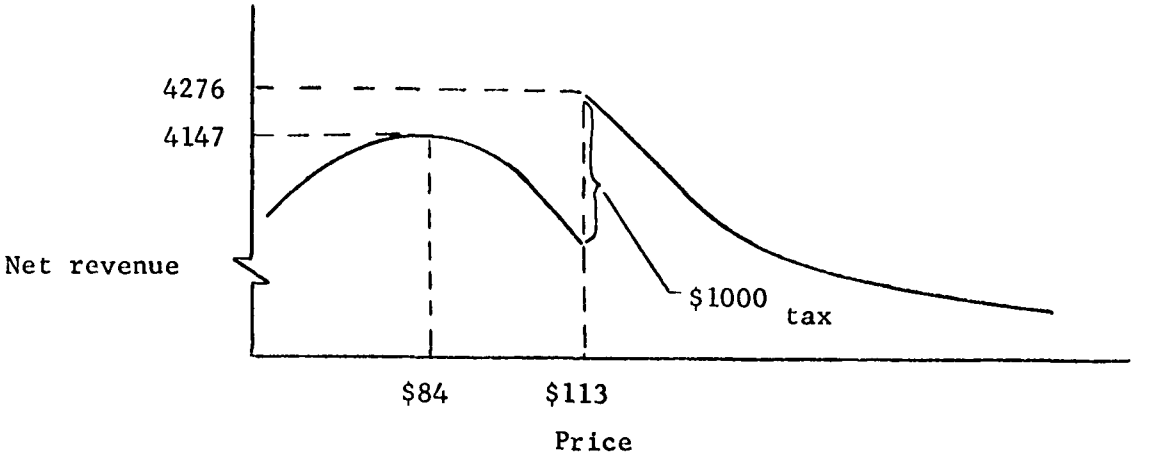


Figure III

Indeed, in this case, forcing the price to be less than $113 will cause the optimum price to be $84 and the optimum import quantity to be 164 units. (Revenue = 4147) This price difference is acceptable, so that the final solution (if the design procedure ends at this point) will be:

Market	1	2	3	4
Variant	11	11	1	2
Price	88	84	160	110
Quantity to Import	132	164	101	115

Total Profit = $18,594

The solution for the first twenty five revenue values took less than ten seconds on a TSS 360/67. Problem B was done by hand in a few minutes for this example.

VI. Conclusion and Discussion

The purpose of any model is to lend insight into the nature of a problem. In this paper, I have been concerned with the problem of determining the optimum product-line expansion policy. Major emphasis has been placed on expansion of modifications of existing products into foreign markets. As in any model, the level of usefulness depends on (1) the exactness with which it quantifies nature, (2) the degree of reliability of the inputs, and (3) an awareness, on the part of the user, concerning the assumptions of the model.

The model presented in this paper is quite general from the standpoint of the functional relationships which can be used for the demand curves. The assumptions of costs, linear in terms of quantities sold, are not drastic; the solution could probably be generalized for other types of cost functions. The importance of the model is, first, that it emphasizes the types of costs and constraints which a manager must face when going international. And, secondly, that it indicates how these considerations interact.

Several drawbacks of the model and their significance must be pointed out at this time:

1.) In order to apply the model a good deal of information on existing or possible product variants must be available. The model becomes one of suboptimization since we can never determine all possible product modifications.

Nevertheless, an interative technique, such as that outlined in the example problem, can give the designer good insights into the optimal direction of modifications.

2.) Certain independencies have been assumed in the model. Allowance for cross elasticity of demand or interaction between markets [12], [23], [38], is not built into the model. Costs of producing different variants are assumed to be independent of each other. Only one adaptation of each "product" is allowed in a market. For air conditioning, for instance, large total facility units must be considered one product, while small single room units are considered a second product. Clearly, if both products are used in a market, the demand for each will be affected.

An iterative solution procedure becomes necessary so that the revenue from the case where both types of air conditioning are sold can be considered. In the example the interaction between prices in Belgium and France, displayed how this iteration would be achieved. This iteration procedure can be used to solve difficulties caused by dependent effects. It is, however, heuristic in nature--not necessarily converging to the optimum.

3.) I have considered a one-period problem which yields results only for that time period. Nonstationary demands are, of course, an important aspect of new market introduction. The existing model can be modified to incorporate this aspect by using total demand over the life of the product with appropriate discount rates for various time intervals. The difficulty of determining life cycles has been addressed in [39],[40].

4.) Demand is a random variable. Using expected demand in the model overlooks the difficulties and inherent risk caused by variance in demand. This difficulty cannot be circumvented readily. One technique would be to predict an average demand, a maximum demand and a minimum demand for each variant in each market. This procedure would give some insight into the range of returns from various strategies.

This paper has presented a model for the product strategy of introducing existing products or minor modifications of existing products into new markets. A sclution procedure is given which first calculates the optimum pricing-inventory strategy for each variant and market. It then searches these single strategies for the optimum overall marketing strategy. An iterative solution procedure is necessary if product or market interaction exists and to take advantage of design insights gained by calculating the optimum strategy. The example problem with four markets and ten modifications took less than ten seconds to solve the first phase on a TSS 360/67. The second phase was done by hand in a few minutes. Time to solve the first phase increases linearly with the number of market-modification strategies available. Time for the second phase will (using Efroymson's results) certainly remain less than 10 minutes. It can be seen that extremely large (200 variants x 50 markets) problems could be solved with this procedure in a very reasonable amount of computer time.

REFERENCES

1. "Air Conditioning Around the World," Air Conditioning, Refrigeration and Heating News, 116, April 14, 1969, (23).

2. Air Conditioning, Refrigeration and Heating News, 110-119, 1966-1967.

3. Alexander, M., "The Significance of Ethnic Groups in Marketing New-Type Packaged Foods in Greater New York," in Advanced Marketing Efficiency, L. Stockman (ed.), American Marketing Association, 1959, (557-561).

4. American Society of Heating, Refrigeration and Air Conditioning Engineers, AHSRAE Guide and Data Book Applications for 1966 and 1967, New York, (1966).

5. Ansoff, H. Igor, Corporate Strategy, (New York: McGraw-Hill Book Company, 1965).

6. Bass, F., D. Tigert, and R. Lonsdale, "Market Segmentation: Group Versus Individual Behavior," Journal of Marketing Research, August 1968, (264-270).

7. Buzzell, Robert D., "Can You Standardize Multi-National Markets?" Harvard Business Review, November 1968, (102-113).

8. Claycamp, H. and W. Massy, "A Theory of Market Segmentation," Journal of Marketing Research, November 1968, (388-394).

9. Day, Ralph L., Marketing Models, (Schanton: International Textbook Company, 1964).

10. Dickter, E., "The World Customer," Harvard Business Review, July-August 1962, (51-60).

11. Efroymson, M. and Ray, T., "A Branch and Bound Algorithm for Plant Location," Operations Research, 14, 1966, (361-368).

12. Fatt, A., "The Danger of 'Local' International Advertising," Journal of Marketing, January 1969, (60-62).

13. Frank, R., "Segmentation Research, Findings and Implications," in Application of the Sciences in Marketing Management, F. Bass, C. King and E. Pessemier (eds.), (New York: John Wiley and Sons, 1963, 39-68).

14. Green, P. and D. Tull, Research for Marketing Decisions, (Englewood Cliffs: Prentice-Hall, 1970).

15. Hall, E., "The Silent Language in Overseas Business," Harvard Business Review, May-June 1960, (425-437).

16. Henderson, J. and R. Quandt, Microeconomic Theory, (New York: McGraw Hill Book Company, 1958).

17. Hillier, F. and Lieberman, G., Introduction to Operations Research, (San Francisco: Holden-Day, Inc., 1967).

18. Hollander, Stanley C., Multi-National Retailing, MSU Intercontinental Business and Economic Studies, 1970.

19. "How the World Imports," International Commerce, November 10, 1969 (4-15).

20. Keegan, W., "Multinational Product Planning: Strategic Alternatives," Journal of Marketing, January 1969, (58-62).

21. Kotler, Phillip, Marketing Management, (Englewood Cliffs: Prentice-Hall, 1967).

22. Kuehn, A. and R. Day, "Strategy of Product Quality," Harvard Business Review, 1962, (100-110).

23. Lenormand, J., "Is Europe Ripe for Integration of Advertising," The International Advertiser, March 1964, (14).

24. Levy, S., "Social Class and Consumer Behavior," in On Knowing the Consumer, J. Neuman (ed.), (New York: John Wiley and Sons, 1966, 146-160).

25. Pessemier, Edgar A., New Product Decisions, (New York: McGraw-Hill Book Company, 1966).

26. Robinson, Richard D., "The Challenge of the Underdeveloped National Market," Journal of Marketing, October 1961, (19-25).

27. _____, International Management, (New York: Holt, Rinehart and Winston, 1967).

28. Rutenberg, David, "Design Commonality to Reduce Multi-Item Inventory: Optimal Depth of a Product Line," Operations Research, 19, 1971 (491-509).

29. _____, "Three Pricing Policies for a Multi-National Company," Management Science, forthcoming.

30. _____, and T. Shaftel, "Product Design: Subassemblies for Multiple Markets," Management Science, forthcoming.

31. Shipman, John R., "International Patent Planning," Harvard Business Review, April 1967, (6-72).

32. Simon, J., "The Effect of Advertising on Liquor Brand Sales," Journal of Marketing Research, August 1969, (301-313).

33. Smith, W., "Product Differentiation and Market Segmentation as Alternative Strategies," Journal of Marketing, July 1956, (3-8).

34. Sommers, Montrose and Kernan, "Why Products Flourish Here, Fizzle There," Columbia Journal of World Business, March 1967, (89-97).

35. Spielberg, Kurt, "Algorithms for the Simple Plant Location Problem with Some Side Conditions," Operations Research, 17, 1969, (85-111).

36. Starr, Martin, "Modular Production-A New Concept," Harvard Business Review, November 1965, (131-142).

37. Steflere, Volney, "Market Structures Studies: New Products for Old Markets and New Markets (Foreign) for Old Products," Application of the Sciences in Marketing Management; Bass, King, and Pessemier (eds.) (New York: John Wiley and Sons, 1968).

38. Urban, G., "A Mathematical Modeling Approach to Product Line Decisions," Journal of Marketing Research, February 1969, (40-47).

39. Vernon, R., "International Investment and International Trade in the Product Cycle," Quarterly Journal of Economics, May 1966, (190-207).

40. Wells, L., "A Product Life Cycle for International Trade," Journal of Marketing, July 1968, (1-8).

Management Sciences Research Report No. 248

A SIMPLEX-LIKE ALGORITHM FOR THE CONTINUOUS MODULAR DESIGN PROBLEM

by

Timothy L. Shaftel
and
Gerald L. Thompson

May 1971

This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract N00014-67-A-0314-0007 NR 047-048 with the U. S. Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

Management Sciences Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

ABSTRACT

This paper derives an efficient solution procedure for solving the continuous version of the Evans modular design problem. The Kuhn-Tucker conditions are used to derive a dual problem which can be solved easily and whose dual variables indicate which equations should be tight. The technique retains a tree-basic solution throughout so that fast solution routines can be employed which are quite similar to those for transportation problems. Because of these analogies, the authors predict the solution of transportation size problems with only moderately increased computer time.

1. INTRODUCTION

The modular design problem was first presented by David Evans [4]. In this problem, parts are to be grouped into a single module, several of which are then used in each application. The objective is to minimize the total cost of parts used:

$$\begin{aligned} &\underset{x,y}{\text{Min}} \quad \sum_{i \in I} c_i x_i' \sum_{j \in J} d_j y_j' \\ &\text{s.t.} \quad x_i' y_j' \geq r_{ij}' \\ &\qquad x_i', y_j', c_i, d_j, r_{ij}' \geq 0 \\ &\qquad x_i', y_j' \quad \text{integers} \end{aligned} \Bigg\} \quad \text{For all } i \text{ and } j$$

where

$I = \{1,2,\ldots,m\}$

$J = \{1,2,\ldots,n\}$

c_i = cost of part i

d_j = demand for application j

r_{ij}' = number of part i units required in application j.

x_i' = the number of part i on the module (decision variable)

y_j' = the number of modules needed in application j (decision variable)

In the previous approaches to this problem listed in the bibliography and also in the present paper, the problem is modified to the continuous modular design problem by dropping the integer requirements on x_i' and y_j' for all i and j; it is believed that a solution to the continuous problem is a necessary prelude to any solution of the integer version.

After the Evans paper, a second paper on modular design was written by A. Charnes and M. Kirby [2]. Both of these papers, proposed solution procedures based on searching the x-y space via specialized search routines. A third paper [6], by A. Passy, modified Charnes and Kirby's procedure by formulating the model as a geometric programming problem. The approach presented in Passy's paper is to move from one system of tight constraints to another until the optimum is found. Although Passy's procedure avoids the relatively slow search procedures of the first two papers, it has three major drawbacks. First, convergence of the procedure was not proved. (The results of the present paper might aid in proving convergence for Passy's procedure.) Second, procedures outlined by Passy to solve the problem of cycling could (and most likely, would) lead to an inordinate number of pivots. And, finally, the complexity of the calculations for each pivot step could require excessive amounts of computer time even if convergence were proved.

It is the intent of the present paper to:

1.) Develop a dual problem (with properties similar to the linear programming dual) from the Kuhn-Tucker optimality conditions.

2.) Use the dual solutions to develop a simplex-like solution algorithm and prove its convergence.

The algorithm presented in this paper is expected to be very efficient. The authors conjecture that a continuous modular design problem could be solved with only slightly more computation time than a transportation problem of the same size. They are presently working on computational tests of this conjecture.

2. DUALITY IN THE MODULAR DESIGN PROBLEM

Evans modified the original problem by making the following substitutions:

$$x_i = c_i \, x_i'$$

$$y_j = d_j y_j'$$

$$r_{ij} = r_{ij}' c_i d_j$$

lle also noted that there exists an infinite number of solutions to the problem since if $\bar{x}$ and $\bar{y}$ are solution vectors, then so are $\bar{x}/\theta$ and $\bar{y}\cdot\theta$ for any $\theta > 0$. We may, therefore, add the restriction $\sum_j y_j = 1$ which singles out a unique member of the class without loss of generality. These transformations lead to the primal problem (P):

$$\min_x \sum_i x_i = g \qquad (1P)$$

$$\text{s.t.} \quad x_i y_j - z_{ij} = r_{ij} \quad \text{for all } i \text{ and } j \qquad (2P)$$

$$\sum_j y_j = 1 \qquad (3P)$$

$$z_{ij}, x_i, y_j, r_{ij} \geq 0 \quad \text{For all } i \text{ and } j \qquad (4P)$$

where z_{ij} is a surplus variable.

Theorem 1 (Evans) The solution to problem (P) exists and is unique.

This theorem was first proved by Evans [4]. The algorithm established in the present paper gives an alternate, constructive proof of the existence of a solution to problem (P).

LEMMA 1. If the requirements matrix, R, has no zero rows or columns, then in any solution, $x_i, y_j > 0$ for all i and j.

Proof. The assertion is obvious since $x_i y_j > 0$ at least once for each i and j.

Let λ_{ij} be associated with constraints (2P) and μ be associated with constraint (3P). Then it is easy to show that the Kuhn-Tucker conditions associated with the primal problem are:

$$\sum_j \lambda_{ij} y_j = 1 \quad \text{for all } i \in I \tag{1C}$$

$$\sum_i x_i \lambda_{ij} = \mu \text{ for all } j \in J \tag{2C}$$

$$z_{ij} \cdot \lambda_{ij} = 0 \qquad \text{For all i and j} \tag{3C}$$

$$\lambda_{ij} \geq 0 \qquad \text{For all i and j} \tag{4C}$$

LEMMA 2. For any pair of feasible solutions to the primal and dual problems, $\mu = g$.

PROOF. Multiply (1C) by x_i, sum over all i, and use (1P) to show:

$$\sum_i \sum_j x_i \lambda_{ij} y_j = \sum_i x_i = g$$

Now multiply (2C) by y_j, sum over all j and use (3P) to show:

$$\sum_i \sum_j x_i \lambda_{ij} y_j = \mu \sum_j y_j = \mu$$

Hence, $\mu = g$.

By analogy with classical linear programming we shall interpret constraints (3C) as "complementary slackness" conditions and insure that they hold by the algorithmic solution techniques we develop. Again by analogy to linear programming we add the objective function $\sum_i \sum_j \lambda_{ij} r_{ij}$ to constraints (1C), (2C), and (4C) to create the dual problem. The remainder of this section is devoted to showing that most of the simplex method solution techniques involving interplay between the primal and dual problems can be carried over to the modular design problem considered here. In later sections we show that they are powerful enough to make possible an efficient simplex-like algorithm for solving the continuous modular design problem.

The dual problem is defined by:

$$\underset{\lambda_{ij}}{\text{Maximize}} \quad \sum_i \sum_j \lambda_{ij} r_{ij} = f \tag{1D}$$

$$\text{s.t.} \quad \sum_j \lambda_{ij} y_j = 1 \quad \text{for all } i \tag{2D}$$

$$\sum_i x_i \lambda_{ij} = g \quad \text{for all } j \tag{3D}$$

$$\lambda_{ij} \geq 0 \quad \text{for all } i \text{ and } j \tag{4D}$$

LEMMA 3. For any pair of primal dual solutions (whether nonnegative or not): $g - f = \sum_i \sum_j \lambda_{ij} \cdot z_{ij}$.

PROOF. Multiply (2P) by λ_{ij}, sum over all i,j and use (1D) to show:

$$\sum_i \sum_j \lambda_{ij} x_i y_j - \sum_i \sum_j \lambda_{ij} z_{ij} = \sum_i \sum_j \lambda_{ij} r_{ij} = f$$

but $\sum_i \sum_j \lambda_{ij} x_i y_j = g$ (see proof of Lemma 2)

hence, $g - f = \sum_i \sum_j \lambda_{ij} z_{ij}$.

LEMMA 4. (Complementary Slackness.) For any pair of primal dual feasible solutions $g = f$ iff $\lambda_{ij} \cdot z_{ij} = 0$ for all i,j.

PROOF. The proof follows directly from Lemma 3 and the fact that λ_{ij} and z_{ij} are non-negative for all i and j.

THEOREM 2 (Duality Theorem). The quantities $\bar{x}_i$, $\bar{y}_j$ for all i and j are a solution to problem (P) iff $\bar{\lambda}_{ij}$ for all i and j are a solution to problem (D) and $g = f$.

PROOF. [The proof of this theorem will be only briefly sketched.] The Arrow-Hurwicz-Uzawa Constraint Qualification [5, p. 102] can be shown to hold for the constraint set of the primal problem. This implies that, at the optimum solution to the primal problem there exists a solution to the Kuhn-Tucker problem [5, pp. 105-106]. The functions defining the constraint set can be shown to be quasi-concave in the non-negative orthant while the objective function is linear. Zangwill [12, p. 43] has shown that in such a case a solution to the Kuhn-Tucker problem can occur only at

the optimum of the primal problem. Finally the K-T problem is made up of the constraint set of the dual problem together with the complementary slackness conditions $\lambda_{ij} \cdot z_{ij} = 0$ which are true if and only if $g = f$ (Lemma 4). This completes an outline for the proof of Theorem 2.

A property of interest, although not used here is that problems (P) and (D) are not mutually-dual. If we form the dual of problem (D) we get the following problem:

$$\underset{\zeta_i, \eta_j}{\text{Min}} \quad \sum_i \zeta_i + g \sum_j \eta_j = g' \qquad (1E)$$

$$\text{s.t.} \quad x_i \eta_j + \zeta_i y_j \geq r_{ij} \quad \text{for all } i \text{ and } j \qquad (2E)$$

where, ζ_i is the dual variable associated with constraint (2D) and η_j is the dual variable associated with constraint (3D).

For fixed x_i and y_j problems (D) and (E) are mutually dual generalized transportation problems. If $\bar{x}_i$ and $\bar{y}_j$ are solutions to (P) then $\bar{\zeta}_i = \bar{x}_i$, $\bar{\eta}_j = 0$ are a feasible solution to (E) from which it easily follows that $g \geq g'$.

3. THE RESTRICTED PROBLEM AND ITS SOLUTION

In this section it is first necessary to define some properties of graphs. For a general discussion of graph theory see Berge [1]. The following description was taken from [9, p.2].

"Let V be a set of n elements called <u>vertices</u> or <u>nodes</u> and let E be a set of (some of the) pairs (u,v) with u,v ϵV. A pair (u,v) is called an <u>edge</u> between u and v, or also between v and u (no direction is implied). Then G = (V,E) is called a <u>graph</u>. A <u>path</u> between u and v in G is a list:

$$u = w_0, w_1, \ldots, w_t = v$$

where, $(w_{j-1}, w_j) \in E$ for $j = 1,\ldots,t$. A path is a cycle if $u = v$ in the above list. A graph is acyclic if it has no cycles. A graph is connected if there is at least one path connecting each pair of distinct nodes. A tree is a connected acyclic graph. Equivalently, a graph is a tree if and only if there is a unique path between each pair of distinct nodes."

In addition to the definitions quoted above we shall need the following. A forest is an acyclic graph. It is easy to show that a forest is the union of trees, that is, a union of connected acyclic graphs.

In the modular design problem we shall consider the graph $G = (V,E)$ defined as follows: The set V of nodes consists of the rows and columns of the requirements matrix R; the set E of edges consists of some of the cells (i,j) of the R matrix.

Suppose G has a cycle

$$\Gamma = \{(s_1,t_1), (s_2,t_2),\ldots,(s_\ell,t_\ell)\}$$

where, $s_p = s_{p+1}$ or $t_p = t_{p+1}$ for $p = 1,2,\ldots,\ell$.

(Note $s_{\ell+1} = s_1$ and $t_{\ell+1} = t_1$) and ℓ is an even number. In each column of the R matrix there are either zero or two cells of the cycle. Then Γ can be written. $\Gamma = \Gamma_1 \cup \Gamma_2$ where:

$$\Gamma_1 = \{(s_1,t_1), (s_3,t_3),\ldots,(s_{\ell-1}, t_{\ell-1})\}$$

$$\Gamma_2 = \{(s_2,t_2), (s_4,t_4),\ldots,(s_\ell,t_\ell)\}$$

DEFINITION. The value $w_{(s,t)}$ of a cycle relative to any element $(s,t) \in \Gamma_2$ is defined to be the ratio

$$w_{(s,t)} = \frac{\prod_{(u,v)\in\Gamma_1} r_{u,v}}{\prod_{(k,p)\in\Gamma_2} r_{k,p}}$$

if all $r_{kp} > 0$ for $(k,p)\in\Gamma_2$; otherwise $w_{(s,t)} = \infty$.

DEFINITION. A nondegenerate problem is one having $w_{(s,t)} \neq 1$ for all cycles Γ and $(s,t) \in \Gamma_2$.

LEMMA 5. A problem with data r_{ij} for all i and j, may be replaced by a nondegenerate problem with perturbed data $r^*_{ij} = r_{ij} + \delta^{i+mj}$ if $r_{ij} \neq 0$ or $r^*_{ij} = 0$ if $r_{ij} = 0$ and where δ can be chosen arbitrarily small.

PROOF. For given m and n there are only a finite number of possible cycles Γ. For such a cycle to have value 1 with the perturbed data we must have

$$\prod_{(i,j)\in\Gamma_1} (r_{ij} + \delta^{i+mj}) - \prod_{(s,t)\in\Gamma_2} (r_{st} + \delta^{s+mt}) = 0.$$

This expression is a polynomial in δ. Moveover there is at least one power of δ that has a non zero coefficient. To show this, let (h,k) be the cell in Γ with smallest k, and given this k the smallest h; suppose $(h,k)\in\Gamma_1$ (a similar proof holds for $(h,k)\in\Gamma_2$). Then there is a term $c\,\delta^{h+mk}$ where $c = \prod_{(i,j)\in\Gamma_1-\{(h,k)\}} r_{ij} \neq 0$ since the value of the cycle is one and all other terms have higher powers of δ. Hence we need only choose δ small and not equal to any of a finite number of zeros of a finite number of polynomials to obtain a nondegenerate perturbed problem close to the original one.

In the rest of this paper, we shall assume that we are dealing with a nondegenerate problem. The techniques for extending the algorithm we shall present to degenerate problems are similar to those for linear programming and will not be discussed.

DEFINITION. Given a feasible solution to problem (P) by a tight constraint in row i we shall mean a cell (i,j) such that $x_i y_j - r_{ij} = z_{ij} = 0$.

LEMMA 6. At an optimum solution of problem (P) there is at least one tight constraint in each row and column of R.

PROOF. Assume the contrary, that we have an optimal solution and for some row u no cell is tight, i.e., $x_u y_j > r_{uj}$ for all j. But then we can decrease x_u while keeping the solution feasible and, therefore, decrease the objective function, which is a contradiction. By the problem symmetry, the same kind of proof is valid for columns of R.

LEMMA 7. If x_i, y_j, and z_{ij} are solutions to problem (P) then the graph with nodes being rows and columns of R and edges being the tight cells (i.e., $z_{ij} = 0$) is a forest.

PROOF. The proof follows directly from Lemma 6 and the fact that we are dealing with only non-degenerate problems, therefore eliminating the possibility of cycles.

We shall now define a forest-restricted problem associated with a given forest, F, to be:

$$\min_{x_i, y_j} \sum_{i \in I} x_i = g \qquad (1H)$$

$$\text{s.t.} \quad \sum_{j \in J} y_j = 1 \qquad (2H)$$

$$x_i y_j = r_{ij}, \quad (i,j) \in F \qquad (3H)$$

$$x_i, y_j \geq 0 \quad \text{for all } i \text{ and } j \qquad (4H)$$

A tree restricted problem is a forest-restricted problem where the forest is made up of a single tree. In the algorithm to be presented in this paper the authors retain a tree basis throughout the procedure. We are, therefore, primarily interested in solutions to tree-restricted problems. The solution to a tree-restricted problem will now be characterized.

For any tree-restricted solution there exists a unique path connecting

any two columns of R [1], [9]. This unique path may be presented as in Figure 1, for the case that column 1 is connected to column q. We have displayed only those columns and rows from R that correspond to the path between column q and column 1. In Figure 1 the cells where r's appear are all tight and rows and columns have been permuted and relabeled to be in the staircase form showed. Certain other cells may be tight in these rows and columns but are not of interest at this time, hence they are not indicated in Figure 1.

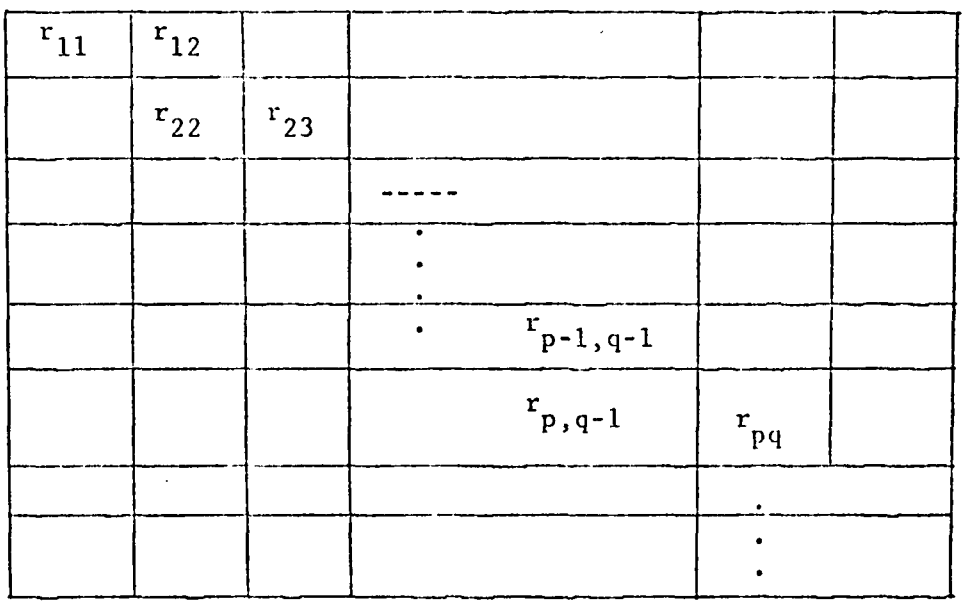


Figure 1

It is easy to see that

$$y_q = \frac{r_{pq}}{r_{p,q-1}} \cdot y_{q-1} \qquad \text{since} \quad x_p y_q = r_{pq}$$

and $x_p y_{q-1} = r_{p,q-1}$. The procedure can be continued in a similar fashion until finally:

$$y_q = \frac{r_{pq}}{r_{p,q-1}} \cdot \frac{r_{p-1,q-1}}{r_{p-1,q-2}} \cdots \frac{r_{23}}{r_{22}} \cdot \frac{r_{12}}{r_{11}} \qquad y_1 = d_{q1} y_1 \quad \text{(1K)}$$

In general denote by d_{ut} the ratio $\frac{y_u}{y_t}$. It will always be the case, as above, that d_{ut} is the quotient of products of r_{ij}'s. We can choose an arbitrary column, say column k, in the matrix R and represent all values of y_j for $j \epsilon J$, in terms of y_k by means of the equation $y_j = d_{jk}\, y_k$. Note that $d_{kk} = 1$. Using constraint (2H) we have the solution to the tree restricted problem as $\sum_{j \epsilon J} y_j = \sum_{j \epsilon J} d_{jk} \cdot y_k = 1$ which yields

$$\therefore \quad y_k = \frac{1}{\sum_{j \epsilon J} d_{jk}}$$

and $y_v = d_{vk} y_k$ for all $v \epsilon J$. The values of x_i are easily determined from constraint (3H).

Given a tree basis T, the associated restricted tree dual solution can be derived by using the fact that $\lambda_{ij} = 0$ for (i,j) not contained in the tree basis. This follows since these cells are not forced to be tight. Also, $z_{ij} = 0$ for (i,j) contained in the tree basis. Taking constraint (2P), multiplying by λ_{ij}/g and summing over i and j separately yields:

$$\sum_i \lambda_{ij}\, x_i\, y_j/g - \sum_i \lambda_{ij} \cdot z_{ij}/g = \sum_i \lambda_{ij} \cdot r_{ij}/g$$

and $$\sum_j \lambda_{ij}\, x_i\, y_j/g - \sum_j \lambda_{ij} \cdot z_{ij}/g = \sum_j \lambda_{ij} \cdot r_{ij}/g$$

using constraints (2D) and (3D) and complementary slackness we may modify these equations to be:

$$y_j = \sum_{(i,j) \epsilon T} \lambda_{ij} \cdot r_{ij}/g \quad \text{for } j \epsilon J \qquad (1L)$$

and $$x_i = \sum_{(i,j) \epsilon T} \lambda_{ij} \cdot r_{ij} \quad \text{for } i \epsilon I \qquad (2L)$$

It is interesting to note that the form of the equations for the variables λ_{ij} are the same as the form derived by Passy [6, p. 450] for the geometric dual

variables if we let $\rho_{ij} = \lambda_{ij} \cdot r_{ij}/g$. The dual objective function developed by Passy, however, is very different from the one used in this paper.

The tree structure of the nonzero variables in equations (1L) and (2L) means that the equations can be solved by a simple solution procedure. Note that they can be rewritten as

$$y_j = \sum_{(i,j)\epsilon T} \rho_{ij} \quad \text{for } j\epsilon J \qquad (1L')$$

$$x_i' = \sum_{(i,j)\epsilon T} \rho_{ij} \quad \text{for } i\epsilon I \qquad (2L')$$

where we have made the substitutions $x_i' = x_i/g$ and $\rho_{ij} = \lambda_{ij} r_{ij}/g$. The following algorithm finds the ρ_{ij}'s given the primal solution x_i', y_j:

(1) Let TR(TC) be the set of rows (columns) containing a unique tight cell. Because we have a tree-restricted solution TR $\cup$ TC is not empty.

(2) For all tight cells (i,j) with $i\epsilon$TR let $\rho_{ij} = x_i'$. For all tight cells (i,j) with $j\epsilon$TC let $\rho_{ij} = y_j$. That this is correct follows from the fact that these tight cells are unique in their rows or columns.

(3) "Cross out" the rows $i\ \epsilon$TR and $j\ \epsilon$TC. For the remaining matrix define a new primal solution $\bar{x}_i'$ and $\bar{y}_j$ as follows

$$\bar{x}'_i = x_i' - \sum_{\substack{j\epsilon TC \\ (i,j)\epsilon T}} y_j \quad \text{for all } i \notin TR$$

$$\bar{y}_j = y_j - \sum_{\substack{i\epsilon TR \\ (i,j)\epsilon T}} x_i' \quad \text{for all } j \notin TC$$

(4) If there are no uncrossed out rows stop; otherwise go back to step (1) and repeat.

A few additional properties of tree-restricted solution will now be presented.

LEMMA 8. For any tree restricted solution let T be the tree basis and let (p,h) be a cell not in T. Then $T \cup \{(p,h)\}$ has a unique cycle whose value is

$$w_{(p,h)} = \frac{x_p y_h}{r_{ph}} .$$

PROOF. Let us assume (without loss of generality) that h = 1 and column 1 is connected to row p as in Figure 1. Adding the cell (p,h) = (p,1) we obtain the cycle shown in Figure 2. We know that $y_q = d_{q1} y_1$ and

	1	2	. . . .		q		
1	r_{11}	r_{12}	....			....	
2	.	r_{22}	....				
.	.	.	....				
.				$r_{p-1,q-1}$			
p	$r_{p,1}$			$r_{p,q-1}$	$r_{p,q}$		
	.	.		.	.	.	

Figure 2

also that $x_p = r_{pq}/y_q$. Hence

$$x_p y_1 = \frac{r_{pq}}{d_{q1}}$$

where, as in (1K),

$$d_{q1} = \frac{r_{pq}}{r_{p,q-1}} \quad \frac{r_{p-1,q-1}}{r_{p-1,q-2}} \quad \cdots \quad \frac{r_{23}}{r_{22}} \cdot \frac{r_{12}}{r_{11}}$$

It follows that

$$\frac{x_p y_1}{r_{p1}} = \frac{r_{pq}}{r_{p1}} \frac{r_{p,c-1}}{r_{pq}} \frac{r_{p-1,q-2}}{r_{p-1,q-1}} \cdots \frac{r_{22}}{r_{23}} \frac{r_{11}}{r_{12}} = w_{(p,1)},$$

as was to be shown.

COROLLARY. A tree-restricted solution x_i, y_j is primal feasible if and only if every non basic cell determines a cycle whose value is ≥ 1.

This corollary is used to check for primal feasibility in the algorithm to be presented.

DEFINITION. Given a tree basis T and any cell $(p,q) \notin T$ we define the following four sets:

I_q = {set of all rows that can be reached in T using cell (p,q), except for row p}

$I_p = I - I_q$

J_p = {set of all columns that can be reached using cell (p,q), except for column q}

$J_q = J - J_p$

Clearly $p \in I_p$ and $q \in J_q$ and these sets are never empty. Also at most one of the sets I_q and J_p are empty.

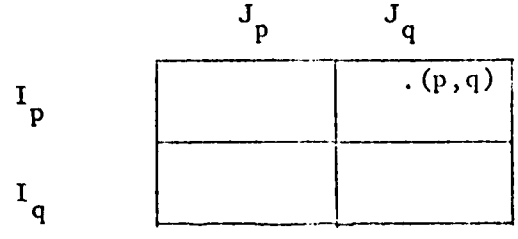


Figure 3

Figure 3 shows the matrix R divided into the four subsets $I_p \times J_p$, $I_p \times J_q$, $I_q \times J_p$, and $I_q \times J_q$. Of these four sets; $I_q \times J_p$ contains no cells of T; $I_p \times J_q$ contains only the cell $(p,q) \in T$; all the rest of the cells of T are in the other two areas $(I_p \times J_p) \cup (I_q \times J_q)$.

In the algorithm to be presented later we are going to change data elements r_{pq} in a parametric fashion. The next two lemmas characterize what happens.

LEMMA 9. Let x_i, y_j be a primal feasible tree-restricted solution with tree basis T. If we replace r_{pq} by a larger value $r^*_{pq} \geq r_{pq}$, then, provided

$$r^*_{pq} \leq \underset{(s,t)\in I_q \times J_p}{\text{Min}} \; r_{pq} w_{(s,t)}$$

the tree-restricted solution x^*_i, y^*_j with the same basis, T, is primal feasible.

PROOF. By the corollary to Lemma 8 we need only show that every non basic cell (u,v) determines a cycle with value $w_{(u,v)} \geq 1$. Referring to Figure 3 it is obvious that non basic cells in the $(I_p \times J_p)$ or $(I_q \times J_q)$ areas have cycles entirely contained in these areas. Hence changing r_{pq} does not affect the values of their cycles.

Because every cycle goes alternately from row nodes to column nodes, every non-basic cell (u,v) in the $I_p \times J_q$ area determines a cycle Γ which includes (p,q) in the Γ_1 and (u,v) in the Γ_2 part. Hence $w_{(u,v)}$ increases if r_{pq} increases and primal feasibility continues to hold for these cells however large we make r^*_{pq}.

Finally consider cells (s,t) in the $I_q \times J_p$ area. Such cells determine a cycle Γ with (s,t) and (p,q) in the Γ_2 part. Let $w_{(s,t)}$ and $w^*_{(s,t)}$ be the value of the cycle determined by (s,t) with r_{pq} and $r^*_{pq} \geq r_{pq}$ respectively. Then $w^*_{(s,t)}\, r^*_{pq} = w_{(s,t)}\, r_{pq}$ by the definition of the value of a cycle. Since we want $w^*_{(s,t)} \geq 1$ this means that we must have $r^*_{pq} \leq w_{(s,t)}\, r_{pq}$ for every $(s,t) \in I_q \times J_p$ and therefore the statement of the Lemma is true.

LEMMA 10. Let x_i, y_j be a primal feasible tree-restricted solution with the tree basis T. If we replace r_{pq} by a smaller value $r^*_{pq} \leq r_{pq}$, then provided

$$r^*_{pq} \geq \max_{(u,v)\in I_p \times J_q - \{(p,q)\}} \frac{r_{pq}}{w_{(u,v)}}$$

the tree-restricted solution x^*_i, y^*_j with the same basis, T, is primal feasible.

PROOF. The proof here is analogous to that for Lemma 9 and will not be given.

4. THE MODIFIED PROBLEM

Given problem R with data r_{ij} we define the modified problem R^* with data r^*_{ij} where $r_{ij} \leq r^*_{ij} < \infty$ for all i and j. We shall give an algorithm for finding an optimal tree-restricted solution to R^* and show how this can be used to find the optimal forest basis solution to Problem R.

THEOREM 3. Given an optimal tree-restricted solution to R^* there corresponds a unique forest-restricted optimal solution to R. Conversely, to the optimal forest-restricted solution to problem R there correspond at least one optimal tree-restricted solution to R^*.

PROOF. Given an optimal tree-restricted solution x^*_i, y^*_j, λ^*_{ij} with tree basis T to problem R^* drop from T all tight cells such that $\lambda^*_{ij} = 0$. The result is a forest F with $\lambda^*_{ij} > 0$ for $(i,j) \in \Gamma$ and $\lambda^*_{ij} = 0$ for $(i,j) \notin F$. Since $x^*_i y^*_j \geq r^*_{ij} \geq r_{ij}$ for all i and j these solutions are primal and dual feasible, and hence by the duality theorem are optimal for problem R.

Given an optimal forest-restricted solution x_i, y_j and λ_{ij} with forest basis F we shall give a constructive procedure for deriving an optimal forest restricted solution to a problem R^*. Suppose $F = T_1 \cup T_2 \cup \ldots \cup T_k$ where each T_i is a tree. If row u contains a (tight) cell of T_i then it will not contain a cell from any other tree. Similarly, if column v contains a cell of T_i then it will not contain a cell of any other tree. We now show one way to "hook together" the trees in F. and make them into a single tree.

Let i_1 be the index of any row containing a cell of T_1 and let $j_2,\ldots,j_k$ be indices of columns containing cells of $T_2,\ldots,T_k$. Add the cells (i_1,j_2), $(i_1,j_3),\ldots,(i_1,j_k)$ to F which will make it into a connected tree and also define problem R^* by

$$r^*_{i,j_2} = x_{i_1} y_{j_2} \geq r_{i,j_2},\ldots,r^*_{i,j_k} = x_{i_1} y_{j_k} \geq r_{i_1 j_k}$$

and all other $r^*_{ij} = r_{ij}$. It follows that x_i, y_j and λ_{ij} are still primal and dual feasible and hence optimal for R^*.

Obviously, there are many other ways the trees $T_1,T_2,\ldots,T_k$ may be connected to make a single tree so the above process is not unique.

THEOREM 4. Let x_i, y_j be a feasible solution to problem R^* with tree basis T and let λ_{ij} be the restricted dual solution; and consider a cell $(p,q) \in T$; then

(a) if $\lambda_{pq} < 0$ we can decrease g^* by increasing r^*_{pq}

(b) if $\lambda_{pq} > 0$ we can decrease g^* by decreasing r^*_{pq}.

These results hold only over a sufficiently small range.

PROOF. Suppose we set $r^*_{ij} = r_{ij} + \delta_{ij}$ and write the Lagrangian function of the primal problem. It is

$$L(x,y,\lambda,\mu,z,\delta) = \sum_i x_i - \sum_i \sum_j \lambda_{ij}[x_i y_j + z_{ij} - (r_{ij} + \delta_{ij})]$$

$$+ \mu\left(\sum_j y_j - 1\right)$$

Now holding all $\delta_{u,v}$ fixed at zero except for δ_{ij} and letting $g(\delta_{ij})$ be the corresponding value of the primal problem we can rewrite the Lagrangian (for small changes in δ_{ij}) as

$$g(\delta_{ij}) = g(0) + \lambda_{ij}\,\delta_{ij} + o(\delta_{ij})$$

for which the two assertions are obvious.

THEOREM 5. Given a feasible tree-restricted solution x_i^*, y_j^*, λ_{ij}^* to problem R^* with basis T, let $(p,q) \in T$. Define r_{pq}^o to be the value of r_{pq}^* for which $\lambda_{pq}^* = 0$, then

$$r_{pq}^o = \sqrt{\frac{\sum_{(i,j)\in R_p} \frac{r_{ij}^*}{e_{ip}}}{\sum_{i\in I_p} e_{ip}} \cdot \frac{\sum_{(i,j)\in C_q} \frac{r_{ij}}{d_{jq}}}{\sum_{i\in J_q} d_{jq}}}$$

where the quantities e_{ip}, d_{jq}, R_p, and C_q will be explained in the proof below. Then

(a) $\lambda_{pq}^* = 0 \Leftrightarrow r_{pq}^* = r_{pq}^o$

(b) $\lambda_{pq}^* < 0 \Leftrightarrow r_{pq}^* < r_{pq}^o$

(c) $\lambda_{pq}^* > 0 \Leftrightarrow r_{pq}^* > r_{pq}^o$

PROOF. Let the sets I_p, I_q, J_p and J_q be as previously defined. If we remove (p,q) from T then as noted before all the remaining cells of T are in the areas $I_p \times J_p$ and $I_q \times J_q$.

For each $j \in J_p$ let i be the smallest row index such that $(i,j) \in T$; and let R_p be the set of such cells (i,j). Then $y_j = r_{ij}/x_i$ for $(i,j) \in R_p$ Also let e_{ip} be the ratio $e_{ip} = x_i/x_p$. Similarly for each i in J_q let j be the smallest column index such that $(i,j) \in T$; and let C_q be the set of all such cells. Then $x_i = r_{ij}/y_j$ for $(i,j) \in C_q$. As before let $d_{jq} = y_j/y_q$.

Removing (p,q) from T forces λ_{pq} to 0 and we can calculate the new solution x_i and y_j for all i and j. From this we can find r_{pq}^o as $x_p y_q$. Once (p,q) is dropped from T the primal problem (P) becomes

$$\text{Minimize} \sum_{i\in I_p} x_i + \sum_{i\in I_q} x_i = \left(\sum_{i\in I_p} e_{ip}\right) x_p + \left(\sum_{(i,j)\in C_q} \frac{r_{ij}}{d_{jq}}\right) \frac{1}{y_q}$$

subject to the constraint

$$\sum_{j\epsilon J_p} y_j + \sum_{j\epsilon J_q} y_j = \left(\sum_{(i,j)\epsilon R_p} \frac{r_{ij}}{e_{ip}} \right) \frac{1}{x_p} + \left(\sum_{j\epsilon J_q} d_{jq} \right) y_q = 1 .$$

Taking the second form of the expression in each case we see we have a constrained minimization problem in the two variables x_p and y_q. The standard Lagrange multiplier technique gives the unique solution

$$x_p = \frac{1}{L} \sqrt{\frac{\sum_{(i,j)\epsilon R_p} \frac{r_{ij}}{e_{ip}}}{\sum_{i\epsilon I_p} e_{ip}}} .$$

and

$$y_q = L \sqrt{\frac{\sum_{(i,j)\epsilon C_q} \frac{r_{ij}}{d_{jq}}}{\sum_{j\epsilon J_q} d_{jq}}}$$

where L is a constant which we never need to determine since we are only interested in the product, $x_p y_q$. Multiplying these two expressions together gives r^o_{pq} as stated in the Lemma. Hence, (a) follows.

For the proof of (b) note that by Theorem 4 if $\lambda^*_{pq} < 0$ we can decrease g by increasing r^*_{pq}. But increasing r^*_{pq} must cause λ^*_{pq} to get closer to 0, since otherwise we could increase r^*_{pq} without bound and continue to decrease g. Since λ^*_{pq} is zero only at r^o_{pq} it follows that $\lambda^*_{pq} < 0$ for $r^*_{pq} < r^o_{pq}$ and $\lambda^*_{pq} > 0$ for $r^*_{pq} > r^o_{pq}$.

THEOREM 6. Let x_i, y_j, λ_{ij} be the optimum tree restricted solution to R^* with basis T. For this optimum solution we have

(a) $\lambda_{pq} > 0 \Leftrightarrow (p,q)\epsilon T$ and $r^*_{pq} = r_{pq} > r^o_{pq}$

(b) $\lambda_{pq} = 0$ and $(p,q)\epsilon T \Leftrightarrow r^*_{pq} = r^o_{pq} \geq r_{pq}$

(c) $\lambda_{pq} = 0$ and $(p,q)\notin T \Leftrightarrow r^*_{pq} = r_{pq}$.

PROOF. (a) If $\lambda_{pq} > 0$ then $r^*_{pq} = r_{pq}$ since $r^*_{pq} > r_{pq}$ would imply that we could reduce g still further. Also $r_{pq} > r^o_{pq}$ by Theorem 5.

(b) If $\lambda_{pq} = 0$ then we can drop (p,q) from T and get the same optimum solution (Theorem 3). By definition then $r^*_{pq} = x_p\, y_q = r^o_{pq}$. Since we have a primal feasible solution $r^*_{pq} \geq r_{pq}$.

(c) This condition is necessary in order that R^* be a nondegenerate problem. In any case there is no reason to make $r^*_{pq} > r_{pq}$ for cells (p,q) not in the basis.

5. THE ALGORITHM

We now present the algorithm for simultaneously solving both problems R^* and R.

(0) Starting Routine. Find a tree restricted primal feasible solution to R^*. (We shall later discuss two ways to find such an initial solution.)

(1) Dual Solution Routine. Find the tree restricted dual solution with tree basis T to R^* by the method outlined in Section 3.

(2) Improvement Routine. Examine the cells (p,q) in the current basis T to see which of conditions (a), (b), or (c) below is true.

(a) There is a cell $(p,q) \in T$ with $\lambda_{pq} < 0$. Calculate r^o_{pq} and raise r^*_{pq} from its current value until either

(i) a cell (h,k) in $I_q \times J_p$ becomes tight; then add (h,k) to T, drop (p,q) from T and set $r^*_{pq} = r_{pq}$, go to 2.

(ii) $r^*_{pq} = r^o_{pq}$, go to 2.

(b) There is a cell $(p,q) \in T$ with $\lambda_{pq} > 0$ and $r^*_{pq} > r_{pq}$. Calculate r^o_{pq} and reduce r^*_{pq} until

(i) $r^*_{pq} = r_{pq}$; go to 2.

(ii) $r^*_{pq} = r^o_{pq}$; go to 2.

(iii) some cell (h,k) in $I_p \times J_q - \{(p,q)\}$ becomes tight; then add (h,k) to T remove (p,q) from T, set $r^*_{pq} = r_{pq}$, and go to 2.

(c) All cells $(p,q) \in T$ satisfy $\lambda_{pq} \geq 0$ and $r^*_{pq} = r_{pq}$ or $\lambda_{pq} = 0$ and $r^*_{pq} = r^o_{pq} \geq r_{pq}$. Go to 4.

(4) The optimum tree-restricted solution to R^* is given by the current T, x_i, y_j and λ_{ij}, since these satisfy the duality theorem. The optimum forest F for R is obtained by dropping those cells (p,q) from T with $\lambda_{pq} = 0$; the same x_i, y_j and λ_{ij} are optimal for R with forest F.

THEOREM 7. For a non-degenerate problem the algorithm converges in a finite number of steps to the optimum answer to problem R.

PROOF. It is obvious that an initial primal forest-restricted feasible solution to R can always be found by choosing an arbitrary set of positive y_j's adding to 1, and then making each x_i just large enough to attain primal feasibility in its row. This solution can then be extended to a feasible tree-restricted solution to problem R^* by following the procedure of Theorem 3.

At each step of the algorithm we move from one primal feasible tree-restricted solution to another one of strictly lower value (because of non-degeneracy). There are only a finite number of tree-restricted bases and the minimum of g is bounded below by 0. Hence we must eventually find a tree-restricted solution that cannot be improved on, i.e. one that satisfies 3(c) of the algorithm. But then the duality theorem is satisfied and we have found the optimum.

It is likely that degeneracy will be even less of a difficulty for this algorithm than it is for the standard transportation problem. But, in any case, similar kinds of degeneracy prevention procedures will work in both cases.

We now discuss two ways of implementing the starting routine of the algorithm. The first procedure is similar to the improvement routine of the algorithm.

Starting Routine 1. Find an initial tree basis T by any means. A good heuristic is to try and get as many of the large entries in R as possible into this initial basis. Now solve for x_i and y_j using T. For each non basic cell (i,j) if $x_i\, y_j < r_{ij}$ replace r_{ij} by $r^*_{ij} = x_i y_j$. After this has been done go back and for each (i,j) such that $r^*_{ij} < r_{ij}$ increase r^*_{ij} until either a new cell becomes tight and enters the basis in place of (i,j) or else $r^*_{ij} = r_{ij}$. Note that this makes g constantly increase and hence these steps are just the reverse of the improvement routine of the algorithm. After a finite number of such steps a primal feasible tree-restricted solution to the original problem will be attained.

Starting Routine 2. Select an arbitrary set of positive y_j's such that $\Sigma\, y_j = 1$. A good choice would be to select

$$y_j = (\Sigma_i\, r_{ij})/(\Sigma_i \Sigma_j\, r_{ij})$$

Now choose $x_i = \operatorname{Max}_j (r_{ij}/y_j)$ and put all tight cells into the basis. If there is a column, say q. with no tight cells, select an arbitrary row, say p, and raise r^*_{pq} to the value $x_p y_q$ and add this cell to the basis also. We now have a primal feasible forest basis which can further be extended to a primal feasible tree basis by using the techniques of the proof of Theorem 3.

As a final remark, we would like to discuss how a primal feasible tree-solution can be used to determine the next tree solution. For the first solution we have $y_j = d_{jk} y_k$ for k fixed and all j. If we now change r^*_{pq} for (p,q) being a tight cell, the values of d_{jk} will be

affected only for those y_j such that (p,q) is a part of the unique path from column j to column k. If r^*_{pq} is changed to r^{**}_{pq} then either

$$d^*_{jk} = d_{jk} \frac{r^*_{pq}}{r^{**}_{pq}} \quad \text{or} \quad d^*_{jk} = d_{jk} \frac{r^{**}_{pq}}{r^*_{pq}}$$

depending on where (p,q) is in the path from column j to column k.

6. EXAMPLES

The first example is designed to demonstrate most of the steps of the algorithm. We start with the data and an initial tree as

(5)	(5)	4
3	(10)	(7)
(4)	3	4

Cells (3,3) and (1,3) are not primal feasible since

$$4 > \frac{7 \cdot 5 \cdot 4}{10 : 5} = 2.8 \quad \text{and} \quad 4 > \frac{5 \cdot 7}{10} = 3.5$$

Following Starting Routine 1 we replace the problem by

(5)	(5)	3.5
3	(10)	(7)
(4)	3	(2.8)

whose value is $g = 51.3$. We now bring cell (3,3) into the basis and can remove any cell in $\Gamma_1 = \{(2,3), (1,2), (3,1)\}$. We choose to remove cell (3,1) and now increase $r^*_{3,3}$ from 2.8, trying to raise it to 4, without causing primal infeasibilities. We succeed and obtain the problem:

(5)	(5)	3.5
3	(10)	(7)
4	3	(4)

whose value is $g = 55.93$. We now bring cell (1,3) into the basis and can remove either cell in $\Gamma_1 = \{(1,2), (2,3)\}$ we choose to remove (2,3) (although later it will become evident that the other choice is better) in order to demonstrate more steps of the algorithm. We raise r^*_{13} to 4 without encountering primal infeasibilities and obtain the problem

(5)	(5)	(4)
3	(10)	7
4	3	(4)

whose value is $g = 56$. Its primal feasible solutions are

$$y = \left(\frac{5}{14}, \frac{5}{14}, \frac{4}{14}\right)$$

$$x = \left(14, 28, 14\right)$$

$$x' = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$$

and the dual solutions are given by

$$\rho_{11} = y_1 = \frac{5}{14}, \quad \rho_{22} = x'_2 = \frac{1}{2}, \quad \rho_{33} = x'_3 = \frac{1}{4}$$

$$\rho_{12} = y_2 - \rho_{22} = \frac{5}{14} - \frac{1}{2} = -\frac{2}{14}, \quad \rho_{13} = y_3 - \rho_{33} = \frac{4}{14} - \frac{1}{4} = \frac{1}{28}$$

Since (1,2) is the only cell with negative dual variable we have $p = 1$, $q = 2$, $I_1 = \{2\}$, $J_2 = \{1,3\}$, $R_2 = \{(1,1),(3,3)\}$, $C_1 = \{(2,2)\}$ and $e_{11} = e_{13} = 1$, $d_{22} = 1$. Hence we can calculate r^o_{12} from Theorem 5 as

$$r^o_{12} = \sqrt{\frac{\frac{5}{1} + \frac{4}{1}}{2} \times \frac{\frac{10}{1}}{1}} = \sqrt{45} = 6.71$$

We also have $I_1 \times J_2 = \{(2,1), (2,3)\}$ as the two cells that may become tight as we increase r_{12}. For them we have:

$$r^*_{12} \leq \frac{5 \cdot 10}{3} = 16.67 \quad \text{from cell (2,1)}$$

and

$$r^*_{12} \leq \frac{10 \cdot 4}{7} = 5.71 \quad \text{from cell (2,3).}$$

The smallest constraint comes from cell (2,3) so we bring it into the basis obtaining the problem:

(5)	5	(4)
3	(10)	(7)
4	3	(4)

whose value is $g = 55.18$. Its solution is

$$y = \left(\frac{35}{103}, \frac{40}{103}, \frac{28}{103} \right)$$

$$x = \left(\frac{103}{7}, \frac{103}{4}, \frac{103}{7} \right)$$

$$x' = \left(\frac{4}{15}, \frac{7}{15}, \frac{4}{15} \right)$$

The dual solution is given by:

$$\rho_{11} = y_1 = \frac{35}{103}, \quad \rho_{22} = y_2 = \frac{40}{103}, \quad \rho_{33} = x'_3 = \frac{4}{15}$$

$$\rho_{23} = x'_2 - \rho_{22} = \frac{7}{15} - \frac{40}{103} = .08, \quad \rho_{13} = x'_1 - \rho_{11} = \frac{4}{15} - \frac{35}{103} = -.07$$

Cell (1,3) is the only one with negative dual variable so $p = 1$, $q = 3$, $I_1 = \{2,3\}$, $J_3 = \{1\}$, $R_3 = \{(1,1)\}$, $C_1 = \{(2,2), (3,3)\}$, and $e_{11} = 1$, $d_{23} = \frac{10}{7}$, $d_{33} = 1$.

Hence we calculate

$$r^o_{13} = \sqrt{5 \times \frac{10 \cdot \frac{7}{10} + \frac{4}{1}}{1 + \frac{10}{7}}} = 4.76$$

We also have $I_1 \times J_3 = \{(2,1), (3,1)\}$ so that the other constraints on r^*_{13} are given by

$$r^*_{13} \leq \frac{5 \cdot 7}{3} = 11.67 \qquad \text{from cell } (2,1)$$

$$r^*_{13} \leq \frac{4 \cdot 5}{4} = 5 \qquad \text{from cell } (3,1)$$

Therefore we increase r^*_{13} to 4.76 obtaining the problem

(3)	5	(4.76)
3	(10)	(7)
4	3	(4)

whose value is $g = 54.83$. The primal and dual solutions are:

$$y = (.29, .41, .30)$$

$$x = (16.56, 24.35, 13.92)$$

$$x' = (.30, .44, .26)$$

$$\rho_{11} = y_1 = .30, \quad \rho_{13} = x'_1 - \rho_{11} = 0, \quad \rho_{22} = y_2 = .41$$

$$\rho_{23} = x'_2 - \rho_{22} = .44 - .41 = .03, \quad \rho_{31} = x'_3 = .26.$$

Since all cells satisfy 3(c) of the algorithm, we have the optimum solution to both R and R^*. We next solve Evans' problem [4], which is given with an initial tree basis:

15	(23)	(44)
(13)	13	0
15	17	(35)
(34)	12	(22)

cells (2,2) and (4,2) are not primal feasible since

$$13 > \frac{13 \cdot 22 \cdot 23}{34 \cdot 44} = 4.4$$

$$12 > \frac{23 \cdot 22}{44} = 11.5$$

The modified primal-feasible problem is:

15	(23)	(44)
(13)	4.4	0
15	17	(35)
(34)	11.5	(22)

We now bring cell (2,2) into the basis and remove cell (2,1). The problem remains primal feasible and we can increase the value of r^*_{22} to 13 while remaining primal feasible.

15	(23)	(44)
13	(13)	0
15	17	(35)
(34)	11.5	(22)

$11.5 = \frac{23 \cdot 22}{44}$ so we bring cell (4,2) into the basis and remove cell (4,3). We then increase r^*_{42} to 12 remaining primal feasible.

Initial primal feasible solution to the original problem:

15	(23)	(44)
13	(13)	0
15	17	(35)
(34)	(12)	22

$y = (.49, .17, .33)$

$x = (132.17, 74.70, 105.13, 68.96)$

$x' = (.35, .20, .28, .18)$

Dual variables

$$\rho_{41} = y_1 = .49$$

$$\rho_{42} = x_4' - \rho_{41} = .18 - .49 = -.31$$

$$\rho_{33} = x_3' = .28$$

$$\rho_{22} = x_2' = .20$$

$$\rho_{13} = y_3' - \rho_{33} = .33 - .28 = .05$$

$$\rho_{12} = x_1' - \rho_{13} = .35 - .05 = .30$$

Only cell (4,2) has a negative dual variable.

Take cell (4,2): $I_q \times J_p = \{(1,1), (2,1), (3,1)\}$

For (1,1) $15 \leq \dfrac{23 \cdot 34}{r_{42}}$ so $r^*_{42} \leq 52.1$

For (2,1) $13 \leq \dfrac{13 \cdot 34}{r_{42}}$ so $r^*_{42} \leq 36.8$

For (3,1) $15 \leq \dfrac{35 \cdot 23 \cdot 34}{44 \cdot r_{42}}$ so $r^*_{42} \leq 41.4$

Therefore $r^*_{42} \leq 36.8$ but $r^o_{42} = 25.18$

Therefore we can raise r_{42} to 25.18

15	(23)	(44)
13	(13)	0
15	17	(35)
(34)	(25.18)	22

$y = (.32, .23, .45)$

$x = (98.00, 55.43, 78.01, 107.33)$

$x' = (.29, .16, .23, .32)$

$g = 338.77$

Dual variables:

$$\rho_{41} = y_1 = .32$$

$$\rho_{42} = x_4' - \rho_{41} = 0.0$$

$$\rho_{33} = x_3' = .23$$

$$\rho_{22} = x_2' = .16$$

$$\rho_{13} = y_3 - \rho_{33} = .45 - .23 = .22$$

$$\rho_{12} = x_1' - \rho_{13} = .29 - .22 = .07$$

Since we satisfy 3(c) of the algorithm, we have the optimal solution to R and R^*.

7. CONCLUSIONS

The algorithm presented in this paper has several advantages:

a.) If we must stop before the optimum solution to the problem is reached, the non-linearity makes g close to the optimum. This solution will be much closer than would be the case in a corresponding similar linear problem.

b.) The method always keeps a primal feasible solution so that one can stop the procedure at any time and have a usable solution.

c.) In order to find the negative restricted dual variable, only the basis tree must be searched. A most negative indicator rule would, therefore, be available at low computational cost. (This is not the case for transportation problem (see [11]).

d.) The search for new limiting cells requires searching only the areas $I_p \times J_q$ or $I_q \times J_p$, and not the whole matrix.

e.) Previous solutions to the primal problem can be used to generate feasible solutions to problems with similar data.

f.) Accuracy is not a problem since a solution to a forest basis can be found independently of any previous operations.

Retention of a tree solution throughout the computation and use of previous solutions makes the steps in the algorithm very similar to those of transportation problems. Srinivasan and Thompson report excellent computational results of 175 x 175 transportation problems in fifteen seconds [11]. The number of pivots for a modular design problem should be of the same order of magnitude as for a transportation problem. This would mean that modular design problems would be solved in only slightly more computer time than comparable transportation problems.

REFERENCES

[1] Berge, C., The Theory of Graphs and Its Applications, John Wiley and Sons, Inc., New York, N. Y., 1962.

[2] Charnes, A., and M. Kirby, "Modular Design, Generalized Inverses and Convex Programming," Operations Research, Vol. 13, pp. 836-847 (1965).

[3] Duffin, J., E. Peterson, and C. Zener, Geometric Programming, John Wiley and Sons, Inc., 1967.

[4] Evans, D., "Modular Design -- Special Case in Nonlinear Programming," Operations Research, Vol. 11, pp. 637-647 (1963).

[5] Mangasarian, O., Nonlinear Programming, McGraw-Hill Book Co., 1969.

[6] Passy, A., "Modular Design: An Application of Structured Geometric Programming," Operations Research, Vol. 18, pp. 441-453 (1970).

[7] Rutenberg, D. and T. Shaftel, "Product Design: Subassemblies for Multiple Markets," Management Science, forthcoming.

[8] Shaftel, T., "An Integer Approach to Modular Design," Operations Research, Vol. 19, pp. 130-134 (1971).

[9] Srinivasan, V. and G. Thompson, "Accelerated Algorithms for Labelling and Relabelling Trees," Management Science Research Report No. 228, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1970.

[10] _____ and _____, "Benefit-Cost Analysis of Coding Techniques for the Primal Transportation Algorithm," Management Science Research Report No. 229, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1970.

[11] _____ and _____, "Duality in the Transportation Model - II," Management Sciences Research Report No. 210, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1970.

[12] Zangwill, W., Nonlinear Programming A Unified Approach, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1969.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Graduate School of Industrial Administration Carnegie-Mellon University	Unclassified
	2b. GROUP Not applicable

3. REPORT TITLE

A SIMPLEX-LIKE ALGORITHM FOR THE CONTINUOUS MODULAR DESIGN PROBLEM

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*

Management Sciences Research Report May 1971

5. AUTHOR(S) *(First name, middle initial, last name)*

Timothy L. Shaftel
Gerald L. Thompson

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1971	30	12

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0314-0007 b. PROJECT NO. NR 047-048 c. d.	Management Sciences Research Report No. 248 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* W.P.# 103-70-1

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Not applicable	Logistics and Mathematical Statistics Br. Office of Naval Research Washington, D. C. 20360

13. ABSTRACT

This paper derives an efficient solution procedure for solving the continuous version of the Evans modular design problem. The Kuhn-Tucker conditions are used to derive a dual problem which can be solved easily and whose dual variables indicate which equations should be tight. The technique retains a tree-basic solution throughout so that fast solution routines can be employed which are quite similar to those for transportation problems. Because of these analogies, the authors predict the solution of transportation size problems with only moderately increased computer time.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Unclassified
Security Classification

A-31406

Management Sciences Research Report No. 254

THE CONTINUOUS MULTIPLE MODULAR DESIGN PROBLEM

by

Timothy L. Shaftel
and
Gerald L. Thompson

July, 1971

This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract N00014-67-A-0314-0007 NR 047-048 with the U. S. Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

Management Sciences Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

ABSTRACT

In the present paper we extend our recent work on the continuous single module design problem to the multiple module case. It is assumed that there is a fixed cost associated with each additional module used to solve the problem. The Kuhn-Tucker conditions characterize local optima among which is a global optimum. Modules are associated with partitions and a special class, guillotine partitions, are characterized. Branch and bound, partial enumeration and heuristic procedures for finding optimum or good guillotine partitions are discussed and illustrated with examples.

1. INTRODUCTION

The one-module modular design problem was first stated by David Evans in [3]. In the problem parts are to be grouped into a single module, several of which can then be used to satisfy (or over satisfy) the requirements for parts in a given application. The objective is to minimize the total cost of the parts used for all applications. The formal statement of the problem is:

$$\underset{x,y}{\text{Minimize}} \quad \sum_{i\in I} c_i x_i \quad \sum_{j\in J} d_j y_j$$

Subject to

$$\left.\begin{array}{l} x_i y_j \geq r_{ij} \\ x_i, y_j, c_i, d_j, r_{ij} \geq 0 \end{array}\right\} \text{ for } i\in I,\ j\in J$$

x_i and y_j are integers

where

$I = \{1,2,\ldots,m\}$

$J = \{1,2,\ldots,n\}$

c_i = cost of part i

d_j = demand for application j (an integer)

r_{ij} = number of part i units required in application j (an integer)

x_i = the number of part i units on the module (a decision variable)

y_j = the number of modules needed for the jth application (a decision variable)

The continuous modular design problem is obtained by dropping the integer requirements on x_i and y_j. A finite simplex-like solution procedure for this problem was given by Shaftel and Thompson in [9]. Several search procedures for the problem were previously given by Evans [3], Charnes and Kirby [1], and, more recently, Passey [6].

Evans in [4] and Rutenberg and Shaftel in [7] have extended the modular design (MD) problem to the case where more than one module can be used. This new problem is called the multiple modular design (MMD) problem. Rutenberg and Shaftel formulated the problem for more than one market with certain other costs and devised a heuristic search procedure for locating good integer solutions. Recently Silverman [10, 11] has presented a search procedure for solving the MMD problem, as formulated by Evans, given a solution to the MD problem.

In the present paper we shall extend the results of our earlier paper [9] to solve the continuous MMD problem. It is felt that solutions to the continuous problem will lead to more efficient solutions of the integer problem. We first sharpen the definition of the MMD problem by relating it to partitions of the MD problem. We then define guillotine partitions and study them in detail. Finally we develop branch and bound and heuristic techniques for finding optimal and good guillotine partitions.

2. THE PRIMAL AND DUAL MMD PROBLEMS

The multiple modular design (MMD) problem for a single market requires that parts be grouped into several modules and various combinations of these modules be used in applications. There is a fixed cost of producing each additional module and certain other costs (such as handling costs) which are a linear function of the number of modules needed in each application.

The MMD has been formally stated [7] as follows:

$$\underset{x,y,p}{\text{Minimize}} \quad \sum_k \Big[\sum_i c_i \underset{\sim}{x}_i^{(k)} \sum_j d_j y_j^{(k)}\Big] + \sum_k F^{(k)} + \sum_k \sum_j \underset{\sim}{b}_j y_j^{(k)} = g$$

Subject to

$$\sum_k \underset{\sim}{x}_i^{(k)} \, \underset{\sim}{y}_j^{(k)} \geq \underset{\sim}{r}_{ij} \quad \text{for } i \in I,\ j \in J$$

$$\underset{\sim}{x}_i^{(k)},\ \underset{\sim}{y}_j^{(k)},\ c_i,\ d_j,\ \underset{\sim}{r}_{ij},\ \underset{\sim}{b}_j \geq 0 \text{ for } i \in I,\ j \in J,\ k \in K$$

$$\underset{\sim}{x}_i^{(k)},\ \underset{\sim}{y}_j^{(k)} \quad \text{integers}$$

where

$$I = \{1,2,\ldots,m\}$$

$$J = \{1,2,\ldots,n\}$$

$$K = \{1,2,\ldots,p\}$$

and p (= the number of modules) is a decision variable;

c_i = cost of part i

d_j = demand for application j

$\underset{\sim}{r}_{ij}$ = number of units of part i required in application j

$\underset{\sim}{b}_j$ = cost of placing a module in application j

$F^{(k)}$ = fixed cost of producing the kth module

$\underset{\sim}{x}_i^{(k)}$ = number of units of part i on module k (a decision variable)

$\underset{\sim}{y}_j^{(k)}$ = number of units of module k needed in application j (a decision variable)

In this paper we shall concentrate on the continuous version of the MMD problem and, therefore will omit the integer restrictions on the x and y variables. For the continuous problem we can make the substitutions

$$x_i^{(k)} = c_i \, \underset{\sim}{x}_i^{(k)},\ y_j^{(k)} = d_j \, \underset{\sim}{y}_j^{(k)}$$

$$r_{ij} = c_i \, d_j \, \underset{\sim}{r}_{ij}, \text{ and } b_j = \underset{\sim}{b}_j / d_j$$

in order to simplify the statement of the problem. Initially we will restrict our attention to the problem where the number, p, of modules has been fixed.

In this case we can drop the fixed charge term in the objective function since it is a constant. The problem then becomes:

$$\underset{x,y}{\text{Minimize}} \quad \sum_k \left[\sum_i x_i^{(k)} \sum_j y_j^{(k)}\right] + \sum_k \sum_j b_j y_j^{(k)} = g_p \tag{1A}$$

subject to

$$\sum_k x_i^{(k)} y_j^{(k)} \geq r_{ij} \tag{2A}$$

$$x_i^{(k)}, y_j^{(k)}, r_{ij}, b_j \geq 0 \tag{3A}$$

Later we will show how to calculate an upper bound on p and give a search procedure for minimizing g_p over all possible values of p. To do this we will have to bring the fixed charge terms back into the objective function.

We shall now derive a simplified primal problem, and then use the Kuhn-Tucker conditions to derive a dual problem similar to the one we found for the single module case in [9].

THEOREM 1. In searching for an optimal solution to problem (A) the linear term in (1A) can be ignored.

PROOF. Let $x^{(k)}, y^{(k)}$ for $k \in K$ be feasible for problem (A). If τ_k is any number > 0 then $x^{(k)}/\tau_k, y_k\tau_k$ for $k \in K$ is also feasible. By setting $\tau_k = \tau < 1$ for all k the second term of objective function for this solution is reduced while the first term remains constant. By letting $\tau \to 0$ the linear term can be made arbitrarily small, so that problem (A), has no minimizing solution with $y^{(k)} > 0$. However the objective function is nonnegative and hence has an infimum M over the constraint set. If we solve problem (A) without the linear term we can then use the above transformation to obtain a solution with value arbitrarily close to M. Thus we are justified in solving the continuous problem without the linear term.

It should be remarked that for the integer problem there are additional constraints of the form $y_j^{(k)} \geq 1$ so that problem (A) has a minimum solution that is attained within the constraint set. In the integer case the linear term cannot be ignored, and Theorem 1 does not hold for the integer problem.

In order to solve the continuous MMD problem we can (as in [3,9]) isolate one of the set of optimum solutions by requiring that $\sum_{j\epsilon J} y_j^{(k)} = 1$ for each k. But we shall be even more explicit concerning the $x_i^{(k)}$ and $y_j^{(k)}$ variables, as follows.

Suppose that in the kth module we choose to make some of the $x_i^{(k)}$ or $y_j^{(k)}$ variables equal to zero. In order to derive a dual problem that reflects this we wish to record these zero constraints explicitly. Define p subsets $I_1,\ldots,I_p$ and $J_1,\ldots,J_p$ such that $I_k \subset I$, $J_k \subset J$ for $k\epsilon K$ and

$$I \times J = (I_1 \times J_1) \cup (I_2 \times J_2) \cup \ldots \cup (I_p \times J_p)$$

Notice that we do not (as yet) require $(I_h \times J_h) \times (I_k \times J_k) = \emptyset$ for $h \neq k$, i.e., we do not require that the sets $I_k \times J_k$ partition $I \times J$. We now call the following problem the primal problem with p modules:

$$\underset{x}{\text{Minimize}} \quad \sum_{k\epsilon K} \sum_{i\epsilon I_k} x_i^{(k)} = g_p \tag{1P}$$

Subject to

$$\sum_{k\epsilon K} x_i^{(k)} y_j^{(k)} - z_{ij} = r_{ij}, \quad i\epsilon I, \ j\epsilon J \tag{2P}$$

$$\sum_{j\epsilon J_k} y_j^{(k)} = 1 \qquad k\epsilon K \tag{3P}$$

$$x_i^{(k)} = 0 \qquad k\epsilon K, \ i\epsilon I - I_k \tag{4P}$$

$$y_j^{(k)} = 0 \qquad k\epsilon K, \ j\epsilon J - J_k \tag{5P}$$

$$z_{ij}, \ x_i^{(k)}, \ y_j^{(k)} \geq 0 \qquad k\epsilon K, \ i\epsilon I, \ j\epsilon J \tag{6P}$$

where z_{ij} is a surplus variable.

THEOREM 2. An optimal solution to problem (P) exists.

PROOF. Problem (P) is to minimize a continuous function over a compact set and by a well-known theorem of mathematics has an optimal solution.

In this paper we shall indicate some algorithms and some heuristic procedures for finding exact and approximate solutions to the continuous MMD problem.

LEMMA 1. If there are no zero rows or zero columns in the requirements matrix, R, then in any feasible solution, for each i there is at least one k with $x_i^{(k)} > 0$, and for each j there is at least one k with $y_j^{(k)} > 0$.

PROOF. Since $\sum_k x_i^{(k)} y_j^{(k)} \geq r_{ij} > 0$ for a feasible solution, it follows that there is at least one k with $x_i^{(k)} y_j^{(k)} > 0$.

Let λ_{ij} be variables associated with the constraints (2P) and let $g^{(k)}$ be variables associated with constraints (3P). The Kuhn-Tucker conditions associated with the primal problem can be shown to be:

$$\sum_{j \epsilon J_k} \lambda_{ij} y_j^{(k)} = 1 \qquad \text{for } k \epsilon K,\ i \epsilon I_k \tag{1B}$$

$$\sum_{i \epsilon I_k} \lambda_{ij} x_i^{(k)} = g^{(k)} \qquad \text{for } k \epsilon K,\ j \epsilon J_k \tag{2B}$$

$$z_{ij} \lambda_{ij} = 0 \qquad \text{for } i \epsilon I,\ j \epsilon J \tag{3B}$$

$$\lambda_{ij} \geq 0 \qquad \text{for } i \epsilon I,\ j \epsilon J \tag{4B}$$

LEMMA 2. For any pair of dual feasible solutions, $\sum_{i \epsilon I_k} x_i^{(k)} = g^{(k)}$ for $k \epsilon K$.

PROOF. Multiply equation (1B) by $x_i^{(k)}$ and sum over $i \epsilon I_k$; obtaining

$$\sum_{i \epsilon I_k} \sum_{j \epsilon J_k} \lambda_{ij} y_j^{(k)} x_i^{(k)} = \sum_{i \epsilon I_k} x_i^{(k)} \quad \text{for } k \epsilon K.$$

Similarly, multiply (2B) by $y_j^{(k)}$ and sum over $j \epsilon J_k$ and use (3P) to obtain

$$\sum_{j\epsilon J_k} \sum_{i\epsilon I_k} \lambda_{ij}\, y_j^{(k)}\, x_i^{(k)} = g^{(k)} \sum_{j\epsilon J_k} y_j^{(k)} = g^{(k)} \quad \text{for } k\epsilon K$$

Combining these two equations gives the desired result.

LEMMA 3. For any pair of dual feasible solutions $\sum_{k\epsilon K} g^{(k)} = g_p$.

PROOF. This follows from Lemma 2 and (1P).

By analogy to the continuous MD problem (see [9]) and classical linear programming, we shall create a dual problem from constraints (1B), (2B), and (4B) and the objective function. The dual problem is:

$$\underset{\lambda_{ij}}{\text{Maximize}} \quad \sum_i \sum_j \lambda_{ij}\, r_{ij} = f_p \,. \tag{1D}$$

Subject to

$$\sum_{j\epsilon J_k} \lambda_{ij}\, y_j^{(k)} = 1 \qquad \text{for } k\epsilon K,\ i\epsilon I_k \tag{2D}$$

$$\sum_{i\epsilon I_k} \lambda_{ij}\, x_i^{(k)} = g^{(k)} \qquad \text{for } k\epsilon K,\ j\epsilon J_k \tag{3D}$$

$$\lambda_{ij} \geq 0 \qquad \text{for } i\epsilon I,\ j\epsilon J \tag{4D}$$

Constraints (3B) can be interpreted as complementary slackness conditions. They will be forced to hold by the solution procedures we shall present.

Note that for each module, the constraints for fixed k have a similar format to that of the dual problem of the continuous MD problem (see [9]) but restricted to the $I_k \times J_k$ area of the matrix R. Many properties of the continuous MD problem will carry over to the continuous MMD problem.

LEMMA 4. For any pair of feasible primal-dual solutions (whether nonnegative or not) we have

$$g_p - f_p = \sum_i \sum_j \lambda_{ij}\, z_{ij}$$

PROOF. Multiply (2P) by λ_{ij}, sum over all i and j and use (1D) to show:

$$\sum_i \sum_j \sum_k \lambda_{ij} x_i^{(k)} y_j^{(k)} - \sum_i \sum_j \lambda_{ij} z_{ij} = \sum_i \sum_j \lambda_{ij} r_{ij} = f_p .$$

As in the proof of Lemma 2 we also have

$$\sum_i \sum_j \sum_k \lambda_{ij} x_i^{(k)} y_j^{(k)} = \sum_k \sum_i x_i^{(k)} = g_p$$

which proves the lemma.

LEMMA 5. (Complementary slackness). For any pair of dual feasible solutions $g_p = f_p$ if and only if $\lambda_{ij} z_{ij} = 0$ for $i \epsilon I$, $j \epsilon J$.

PROOF. The proof follows directly from Lemma 4 and the fact that both λ_{ij} and z_{ij} are nonnegative.

THEOREM 3 (Duality Theorem). A necessary condition that $x_{i.}^{(k)}$, $y_j^{(k)}$ be optimal for problem (P) is that there exists a solution λ_{ij} to problem (D) and that $g_p = f_p$.

PROOF. Using Lemma 1, we can show that the Arrow-Hurwicz-Uzawa constraint qualification [5, p. 102] holds for problem (P). Therefore at the optimum solution to problem (P) there exists a solution to the Kuhn-Tucker problem, since problem (D) together with $\lambda_{ij} z_{ij} = 0$ (which is true if and only if $g_p = f_p$ from Lemma 5) make up the Kuhn-Tucker conditions. These conditions must hold at the optimum because of the constraint qualification [5, pp. 105-106].

The MD problem had a unique solution so the duality theorem for that case was stronger than here. In the MMD problem there are many pairs of solutions satisfying the conditions of Theorem 5, and these correspond to <u>local optima</u> of g_p. Among these, of course, is a global optimum.

3. CHARACTERIZATION OF LOCAL OPTIMA

In the present section we shall give some properties of local optima for the case of a fixed number, p, of modules. It will be shown that a local optimum to the problem can be found by solving p MD problems.

DEFINITION. By a partition of a MMD problem with data matrix R (hereafter called problem R) into p problems with data matrices $R^{(k)}$ (hereafter called problem $R^{(k)}$) for $k \epsilon K$ we mean $R = \sum_{k \epsilon K} R^{(k)}$, i.e.

$$r_{ij} = \sum_{k \epsilon K} r_{ij}^{(k)} \text{ where } r_{ij}^{(k)} \geq 0.$$

An integral partition is one that also satisfies,

$$r_{ij}^{(k)} = c_i \, d_j \, \underset{\sim}{r}_{ij}^{(k)} \text{ where } \underset{\sim}{r}_{ij}^{(k)} \geq 0 \text{ is an integer}$$

We shall regard each $R^{(k)}$ as defining a MD problem with variables $x_i^{(k)}$, $y_j^{(k)}$ and $\lambda_{ij}^{(k)}$.

THEOREM 4. (A) Given an optimal primal dual solution $x^{(k)}$, $y^{(k)}$ and λ_{ij} to problem R a partition $R = \Sigma R^{(k)}$ can be defined in such a way that $x^{(k)}$, $y^{(k)}$ and λ_{ij} are optimal primal-dual solutions to the MD problem $R^{(k)}$.

(B) Given a partition $R = \Sigma R^{(k)}$, let $x^{(k)}$ and $y^{(k)}$ be optimal primal solutions for the MD problem $R^{(k)}$. Then, if it is possible to choose $\lambda_{ij}^{(k)} = \lambda_{ij}$ for all $k \epsilon K$ as optimal dual solutions to $R^{(k)}$ for all $k \epsilon K$, then $x_i^{(k)}$ and $y_j^{(k)}$ give a local optimum for problem R.

(C) If $g^{(k)}$ is the value of problem $R^{(k)}$ and the condition of (B) holds then

$$g_p = \sum_{k \epsilon K} g^{(k)} .$$

PROOF. (A) From (2P) we have

$$\sum_k x_i^{(k)} \, y_j^{(k)} \geq r_{ij}$$

hence we can choose $r_{ij}^{(k)} \leq x_i^{(k)} \, y_j^{(k)}$ so that $\sum_{k=1}^{p} r_{ij}^{(k)} = r_{ij}$ for each i and j in at least one, and perhaps many different ways. If $z_{ij} = 0$ then $\sum_k x_i^{(k)} \, y_j^{(k)} = \sum_k r_{ij}^{(k)} = r_{ij}$ so $z_{ij}^{(k)} = 0$ for problem $R^{(k)}$. Hence $\lambda_{ij} > 0$ implies $z_{ij}^{(k)} = 0$ and complementary slackness holds. The primal and dual

conditions for problem $R^{(k)}$ are as follows:

(1) $x_i^{(k)} y_j^{(k)} \geq r_{ij}^{(k)}$ for $i \epsilon I_k$, $j \epsilon J_k$

(2) $\sum_{j \epsilon J_k} y_j^{(k)} = 1$ for $k \epsilon K$

(3) $\sum_{i \epsilon I_k} x_i^{(k)} = g^{(k)}$ for $k \epsilon K$

(4) $\sum_{j \epsilon J_k} \lambda_{ij} y_j^{(k)} = 1$ for $k \epsilon K$, $i \epsilon I_k$

(5) $\sum_{i \epsilon I_k} \lambda_{ij} x_i^{(k)} = g^{(k)}$ for $k \epsilon K$, $j \epsilon J$

(6) $\sum_{i \epsilon I_k} \sum_{j \epsilon J_k} \lambda_{ij} r_{ij}^{(k)} = f^{(k)}$ for $k \epsilon K$

(7) $x_i^{(k)}, y_j^{(k)}, \lambda_{ij} \geq 0$ for $k \epsilon K$, $i \epsilon I_k$, $j \epsilon J_k$

We have already shown that (1) holds. Equation (2) is the same as (3P); (3) and (6) are true by definition; (4) is (2D); (5) is (3D); and (7) is included in (4P) and (4D). Hence $x_i^{(k)}$, $y_j^{(k)}$ and λ_{ij} are primal-dual solutions to problem $R^{(k)}$.

(B) If for a partition $\Sigma R^{(k)} = R$ we can choose $\lambda_{ij}^{(k)} = \lambda_{ij}$ for all i and j, then the steps in the proof of (A) are reversible. We omit the rest of the details.

(C) If we sum (6) over k and use the duality theorem for problem $R^{(k)}$ and the partition formula we have

$$\sum_k f^{(k)} = \sum_{k \epsilon K} \sum_{i \epsilon I_k} \sum_{j \epsilon J_k} \lambda_{ij} r_{ij}^{(k)} = \sum_{k \epsilon K} g^{(k)} = g_p = f_p$$

completing the proof of the theorem.

By means of this theorem we see that one way of finding local optima to the MMD problem is to partition R into MD problems $R^{(k)}$ and then use the algorithm of [9] to solve $R^{(k)}$. If it happens that the same dual variables

can be used for each problem $R^{(k)}$ then we have found a local optimum to the MMD problem. If we can enumerate all local optima, we can choose the smallest as the global optimum. There are, of course, an infinite number of arbitrary partitions. We shall concentrate on integral partitions, since there are only a finite number of them.

DEFINITION. By a guillotine partition with p pieces of R we shall mean a partition of $I \times J$ into p "rectangular" subsets

$$I \times J = (I_1 \times J_1) \cup \ldots \cup (I_p \times J_p)$$

where

$$(I_h \times J_h) \cap (I_k \times J_k) = \emptyset \text{ for } h \neq k$$

and a partition $R = \sum_{k \in K} R^{(k)}$ where

$$r_{ij}^{(k)} = r_{ij} \text{ if } (i,j) \in I_k \times J_k$$

$$r_{ij}^{(k)} = 0 \text{ if } (i,j) \in I_h \times J_h \text{ and } h \neq k.$$

By a guillotine row partition we mean a guillotine partition with $J_k = J$ for $k \in K$. Similarly by a guillotine column partition we mean a guillotine partition with $I_k = I$ for $k \in K$.

Clearly guillotine partitions are integral and there are only a finite number of them. Also any guillotine partition can be constructed stepwise by row and-or column guillotine partitions applied alternately to a series of sub-problems.

THEOREM 5. Suppose problem R is divided into p rectangular pieces $R^{(k)}$ by a guillotine partition. If $x_i^{(k)}$, $y_j^{(k)}$ and $\lambda_{ij}^{(k)}$ are primal-dual optimal solutions for $R^{(k)}$ then $x_i^{(k)}$, $y_j^{(k)}$ and $\lambda_{ij} = \sum_{k=1}^{p} \lambda_{ij}^{(k)}$ give a local optimum to problem R.

PROOF. Suppose $R^{(k)}$ consists of the matrix $r_{ij}^{(k)}$ defined by

$$r_{ij}^{(k)} = r_{ij} \quad \text{for } (i,j) \in I_k \times J_k$$

$$r_{ij}^{(k)} = 0 \quad \text{for } (i,j) \notin I_k \times J_k$$

where $I_k \subset I$ and $J_k \subset J$ define the rectangular piece $R^{(k)}$. Then if $i \notin I_k$ we have $x_i^{(k)} = 0$ and if $j \notin J_k$ we have $y_j^{(k)} = 0$. Hence $x_i^{(k)} = 0$ if $(i,j) \notin I_k \times J_k$. Therefore $\lambda_{ij}^{(k)} = 0$ if $(k,j) \notin I_k \times J_k$. It follows that if we define $\lambda_{ij} = \sum_{k=1}^{p} \lambda_{ij}^{(k)}$ it will satisfy the dual conditions (2B)-(4B). Hence by Theorem 4 $x_i^{(k)}$ and $y_j^{(k)}$ give a local optimum for problem R.

LEMMA 6. Let $I_k \times J_k$ be a rectangular piece of a guillotine partition and let $g^{(k)}$ be the optimal value of $R^{(k)}$ considered as an MD problem. Then $g^{(k)} \geq \sum_{(i,j) \in I_k \times J_k} r_{ij}$.

PROOF. Since $\lambda_{ij}^{(k)} = 1$ for $(i,j) \in I_k \times J_k$ is always dual-feasible and $g^{(k)} = f^{(k)}$ by the duality theorem, the result is obvious.

DEFINITION. A sub problem $R^{(k)}$ defined on $I_k \times J_k$ is <u>tight</u> if $g^{(k)} = \sum_{(i,j) \in I_k \times J_k} r_{ij}$.

THEOREM 6. Consider the MMD problem with all fixed costs equal to zero.

(a) For guillotine partition $g_p \geq g_{p+1}$.

(b) Unless all pieces of a guillotine partition with p pieces are tight, there is an additional row or column guillotine partition that can be made such that $g_p > g_{p+1}$.

PROOF. (a) The guillotine partition with p+1 pieces is obtained from that with p pieces by applying a row or column guillotine partition to one of the original pieces. Suppose $I_1 \times J_1$ is to be divided into $(I_{11} \times J_1) \cup (I_{12} \times J_1)$ by a row partition (the proof for column partition

is similar). Then the y vector that was optimal for $I_1 \times J_1$ together with the I_{11} components of x will be feasible for $I_{11} \times J_1$. Similarly the same y vector and the I_{12} components of x will be feasible for $I_{12} \times J_1$. Hence (a) follows.

(b) If the $I_1 \times J_1$ piece for the guillotine partition with p pieces is not tight then some cell $(i,j) \in I_1 \times J_1$ has $z_{ij} > 0$. By making the next partition isolate either the row or the column containing (i,j) we can assure that $g_p > g_{p+1}$.

We give next an example of a problem for which the optimum partition is not a guillotine partition. Consider the problem

$R =$

1	2	3
2	3	4
3	4	5

and suppose we require $p = 2$. The sum of the elements in R is 27, and the reader may easily verify that every guillotine partition has $g_2 > 27$. However the following, non guillotine, partition achieves $g_2 = 27$:

$R =$

1	2	3
2	3	4
3	4	5

$=$

0	1	2
0	1	2
0	1	2

$+$

1	1	1
2	2	2
3	3	3

$$= R^{(1)} + R^{(2)}$$

The optimum solution to $R^{(1)}$ is

$$x = (3, 3, 3), \quad y = (0, \tfrac{1}{3}, \tfrac{2}{3}),$$

and the optimum solution to $R^{(2)}$ is

$$x = (3, 6, 9), \quad y = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3})$$

One optimum dual solution is $\lambda_{ij} = 1$ for all i and j, and there are others.

It is interesting to note that the vertical guillotine partition that defines column 1 as one subproblem and columns 2 and 3 as the other subproblem has a value $g_2 = 27.58$ which is only slightly more than the optimum of 27. We have generally found that guillotine partitions will give local optimum values close to the optimum when they do not themselves provide the optimum.

In the rest of this paper we shall restrict ourselves to guillotine partitions. Other reasons why they are advantageous is that row guillotine partitions restrict the number of different parts that appear on a given module, while column guillotine partitions restrict the number of different modules that appear in a given application. Both of these conditions are probably desirable from a manufacturing point of view, even though they are not specifically included in the objective function.

We intend to take up the question of non-guillotine partitions in another paper.

4. SOLUTION PROCEDURES FOR GUILLOTINE PARTITIONS

Searching successive MD problems of a MMD problem is made easier by the ability to use a type of parametric programming. Given the optimal solution to a MD problem, we can readily determine a good initial feasible solution for any subproblem formed via a guillotine partition, as follows. Let x_i, $i \in I$, y_j, $j \in J$ be the optimal solution to a particular $m \times n$ MD problem. Now assume that we partition the problem into two subproblems made of columns of the original problem (the same reasoning would apply for rows). Let

$$J_k = \{\text{the set of columns which are positive in module } k\}, \quad k = 1 \text{ or } 2.$$

And $$y_j^{(k)} = y_j / \sum_{j \epsilon J_k} y_j \qquad j \epsilon J_k \text{ and } k = 1 \text{ or } 2$$

$$x_i^{(k)} = \max_{j \epsilon J_k} r_{ij}/y_j^{(k)}$$

Let the cells which determine $x_i^{(k)}$ be tight and by increasing the right hand sides of other constraints force these cells to enter the basis until a tree basic solution is formed with $r_{ij}^{(k)*} = x_i^{(k)}\ y_j^{(k)}$ for (i,j) in the tree.

THEOREM 7. (A) The solution described above is primal feasible.

(B) If m and n are > 1 and the original problem is nondegenerate the sum of the objective functions for the initial primal feasible solutions to the subproblems is less than the optimal solution for the original problem.

PROOF. (A) Let $r_{iT}/y_T^{(k)} = \max_{j \epsilon J_k} r_{ij}/y_j^{(k)}$

then, $x_i^{(k)}\ y_j^{(k)} = r_{iT}/y_T^{(k)} \cdot y_j^{(k)} \geq r_{ij}/y_j^{(k)} \cdot y_j^{(k)} = r_{ij}$ for all i,j and k.

(B) Let g be the optimal solution to the original problem. And h_1 and h_2 be the objective functions associated with the initial primal feasible solutions of the two subproblems. Then,

$g = \sum_{i \epsilon I} \sum_{j \epsilon J} x_i\ y_j$. We also know that $x_i = \max_{j \epsilon J} r_{ij}/y_j$ for all i,

so that $g = \sum_{i \epsilon I} \sum_{j \epsilon J} y_j \cdot \max_{j \epsilon J} r_{ij}/y_j$.

Now, $$x_i^{(k)}\ y_j^{(k)} = \frac{y_j}{\sum_{j \epsilon J_k} y_j} \cdot \max_{j \epsilon J_k} r_{ij}/y_j^{(k)}$$

$$= \frac{y_j}{\sum_{j \epsilon J_k} y_j} \cdot \sum_{j \epsilon J_k} y_j \max_{j \epsilon J_k} r_{ij}/y_j$$

$$= y_j \max_{j \epsilon J_k} r_{ij}/y_j$$

And, $$h_1 + h_2 = \left[\sum_{i \epsilon I} \sum_{j \epsilon J_1} y_j \max_{j \epsilon J_1} r_{ij}/y_j + \sum_{j \epsilon J_2} y_j \max_{j \epsilon J_2} r_{ij}/y_j \right] .$$

But for m and $n > 1$ (and a nondegenerate problem) there is at least one row in R which contains only one tight cell so that

$$\max_{j \epsilon J} r_{ij}/y_j > \max_{j \epsilon J_k} r_{ij}/y_j .$$

Hence, $g > h_1 + h_2$.

A solution for a subproblem formed in the above manner is usually a very good initial primal feasible solution. It is, of course, an upper bound on the value of the objective function for that subproblem. A lower bound to the objective function value of a subproblem is $\sum_{i \epsilon I_k} \sum_{j \epsilon J_k} r_{ij}^{(k)}$.

It is important to note that the upper bound calculation is the first (and, probably, most time consuming) step of calculating the exact optimum for a subproblem. Because of this, it seems likely that for any branch and bound enumeration procedure it will pay to calculate exact solutions rather than use approximate lower bounds. It is possible to enumerate, explicitly or implicitly, all possible ways of partitioning the problem into p modules (bounds on the value of p will be discussed in the next section). The organization of such an enumeration, not the calculation of lower bounds, is what would make such a complete enumeration difficult. Because of these difficulties we will present some properties which sould make useful tools for obtaining good heuristic solutions.

Given a solution to a particular MD problem, the greatest value by which we can reduce the value of the objective function is $\sum_i \sum_j z_{ij}$. If we form a partition with a subproblem with a single positive column, then the objective function

will reduce by at least $\sum_j z_{ij}$ for that row. Since m is generally larger than n, vertical partitions seem the most appropriate. Another rationale for using only vertical partitions can be made by counting the number of new tight cells which would be caused by vertical versus horizontal partitioning. Once a vertical partition is made, subdividing the new subproblems into more vertical partitions is even more appropriate. If a total of n partitions were to be made, n modules each satisfy the needs of exactly one application would yield the optimum value of $\sum_i \sum_j r_{ij}$.

Given the value of p, or an estimate for that value, and using vertical partitioning only, it becomes necessary to intelligently group applications to be satisfied by certain modules. (Note that the techniques discussed here will also apply to horizontal partitioning.) This fact would mean that we would want to choose applications to group whose parts requirements have approximately the same ratios. We would also like to group applications, with requirements which are difficult to fulfill, into the smallest groups. Three possible measures which may be used are:

1). The variance (or absolute difference) along each column is an important measure. Since we are dealing with ratios, we would want a measure of the variance of the normalized row entries. The measure of this type that we shall use is

$$\sum_i \left| r_{ij} / \sum_i r_{ij} - \frac{1}{m} \right| .$$

2). The above value would be useless without some sort of indication of the size order of the entries of a column. Ideally, of course, we would compare each pair of columns to see how closely they fit. An easier, but less informative, method is to compare each column with a given standard order.

The sum of the absolute difference between the order and the standard order will be used as a measure.

3). Given that we make only vertical partitions the number of columns in a given module is likely to be on the order of magnitude of n/p. In order to determine which columns should be in the smaller partitions we look at $\sum_i z_{ij}$ for the single module solution. Columns with larger values of $\sum_i z_{ij}$ would be in partitions which satisfy the least number of applications.

We consider the effectiveness of the three heuristics for a specific example in Section 6.

5. DETERMINING THE NUMBER OF MODULES

As mentioned earlier, n is an upper bound on the value of p while 1 is a lower bound. The solutions for these two cases are quite easy to find; the difference between the two solutions (excluding fixed costs) is exactly $\sum_i \sum_j z_{ij} = Z^{(1)}$ where z_{ij} is the surplus for the case where p = 1. Reduction in the value of $Z^{(k)}$ as p increases seems to decrease at least linearly. As more experience in solving MMD problems is obtained, good estimates on p should become available. At present a good estimate (assuming $F^{(p)}$ is an increasing function of p) is the largest value of p such that $F^{(p)} \leq \frac{Z^{(1)}}{2^{p-1}}$. During a search for the exact value of p, an upper bound which is often better than n, as well as a means for eliminating certain possible values between 1 and n can be achieved. These two procedures are outlined in the next two theorems.

THEOREM 8: Let $Z^{(q)} = \sum_i \sum_j z_{ij}$ when q modules are produced. Then an upper bound on p is the largest integer such that:

$$Z^{(q)} - \sum_{k=q+1}^{p} F^{(k)} > 0$$

where $F^{(k)}$ is the fixed cost of the kth additional module.

PROOF. By definition $Z^{(q)}$ is the maximum we can reduce the value of the objective function by producing additional modules. As soon as the total cost of additional modules surpasses this amount, adding a module cannot be profitable.

THEOREM 9: Given $Z^{(q)}$ and θ, the value of the best minimum solution found so far, calculate c as the largest integer such that

$$\sum_i \sum_j r_{ij} + Z^{(q)} + \sum_{k=1}^{c} F^{(k)} < \theta .$$

Then any value of k from (and including) c to q cannot be the optimum.

PROOF. Since $Z^{(q)}$ is a nonincreasing function of q, it is a lower bound on the value of $Z^{(c)}$, $c \leq q$. This implies that a lower bound on the value of the objective function for any value of $c \leq q$ is: $\sum_i \sum_j r_{ij} + Z^{(q)} + \sum_{k=1}^{c} F^{k}$. We need only consider values of c which have a lower bound less than the optimum found thus far. Hence the theorem.

6. EXAMPLE

We will give an example problem, enumerate all possible guillotine partitions, and indicate the values of the heuristic measures defined in section 4. The problem we will use is symmetrical so that only vertical partitions need be considered.

$R =$

1	2	3	1	4
2	9	6	12	10
3	6	16	20	25
1	12	20	25	23
4	10	25	23	36

$\sum_j r_{ij} =$ 11 39 70 81 98

$\sum_i \sum_j r_{ij} = 299$

For $p = 1$ $g^{(1)} = g_1^{(1)} = 367.36.$ $Z^{(1)} = 68.36.$ For $p = 2$ there are 15 possible vertical guillotine partitions. In the following table 1,2,-3,4,5 means subproblem one has r_{ij} for columns 1 and 2 and zeros in columns 3, 4, and 5, while subproblem 2 has columns 3, 4, and 5 at the value, r_{ij}, and zero's in columns 1 and 2.

Partition	Sub-module 1		Sub-module 2		Total	
	Optimum	Excess	Optimum	Excess	Optimum	Excess
1 - 2,3,4,5	11	0	343.0	55	354.0	55
2 - 1,3,4,5	39	0	299.1	39.1	338.1	39.1
3 - 1,2,4,5	70	0	285.1	56.1	355.1	56.1
4 - 1,2,3,5	81	0	266.9	48.9	347.9	48.9
5 - 1,2,3,4	98	0	247.0	46.0	345.0	46.0
1,2 - 3,4,5	57.4	7.4	276.6	27.6	334.0	35.0
1,3 - 2,4,5	91.1	10.1	262.5	44.5	353.6	54.6
1,4 - 2,3,5	100.7	8.7	245.0	37.9	345.7	46.7
1,5 - 2,3,4	118.8	9.8	226.0	36.0	344.8	45.8
2,3 - 1,4,5	126.9	17.9	221.9	31.9	348.8	49.8
2,4 - 1,3,5	134.6	14.6	198.5	19.5	333.1	34.1
2,5 - 1,3,4	160.2	23.2	185.4	23.4	245.6	46.6
3,4 - 1,2,5	165.6	14.6	179.9	31.9	345.5	46.5
3,5 - 1,2,4	176.4	8.4	153.0	22.0	329.4	30.4
4,5 - 1,2,3	200.3	21.3	145.4	25.4	345.7	46.7

For $p = 2$ then the optimum partition is 3,5 - 1,2,4 $g^{(2)} = 329.4,$ $Z^2 = 30.4.$ From the enumeration of $p = 2$ we can enumerate all possible solutions for $p = 3$ and $p = 4.$ For $p = 3$ the best partitions are 1,2 - 3,5-4, $g^{(3)} = 314.4,$ $Z^{(3)} = 15.4.$ For $p = 4$ the best partitions are 1,2 - 3-4-5, $g^{(4)} = 306.4,$ $Z^{(4)} = 7.4.$ Note that the approximation for the reduction in $Z^{(k)}$ given in the last section seems to be very accurate.

Finally, we will show the heuristic measures discussed in section 4.

z_{ij} for the single module case:

0	1	1.2	4	2
1	0	6.5	3	8
1.2	6.5	1.4	.8	0
4	3	.8	0	7
2	8	0	7	0
Sum	8.2	18.5	9.9	14.8	17

Normalized value of r_{ij}

.091	.051	.043	.012	.041
.182	.231	.086	.148	.102
.273	.154	.229	.247	.225
.091	.308	.286	.309	.235
.364	.256	.357	.284	.368
Ave.	.2	.2	.2	.2	.2
Sum of the absolute difference from the mean	.473	.528	.543	.480	.515

Order (ties are given the average value of their possible positions):

$1\frac{1}{2}$	1	1	1	1
3	3	2	2	2
4	2	3	3	4
$1\frac{1}{2}$	5	4	5	3
5	4	5	4	5

Calculating the sum of the absolute value of the differences between pairs of columns we get

pair	value	pair	value
1,2	7	3,4	2
1,3	5	3,5	2
1,4	7	4,5	4
1,5	3		
2,3	4		
2,4	2		
2,5	6		

Note that the optimum solution for $p = 2$ is 3,5 - 1,2,4 where two rows which are similar along the order and variance measures (3 and 5) with large z_{ij}'s have been grouped into a smaller module. Although these measures are not perfect, they do give an a priori indication of which partitions are reasonable.

As a second example, we will solve the problem presented by Evans in [3], for p = 2.

15	23	44
13	13	0
15	17	35
34	12	22
Sum 77	65	101

For this problem Evans in [4] gives a solution of 270. Silverman [11] gives an optimal solution of 269.06 via a search technique. (Both solutions are for non-guillotine partitions.) We shall show one possible vertical guillotine partion, chosen from the magnitude of $\sum_i z_{ij}$ from the single module solution;

these values are:

column	1	2	3
$\sum_i z_{ij}$	30.3	14.5	51.0

We shall solve the partitions 1, 2 - 3. In this case $g_1^2 = 166.9$, $g_2^2 = 101$, $g^2 = 267.9$. This solution was found in a few minutes by hand via the Shaftel-Thompson algorithm applied to the subproblems.

7. CONCLUSIONS

In this paper we have extended the MD solution of [9] to the MMD problem. We have characterized the local optimum and identified a useful subset of these optimum as elements of a, possibly, exhaustive search. In solving the MMD problem by the methods outlined in this paper several advantages are available, as follows.

(1). The solution is based on solving the MD problem which can be solved quickly. Use of previous solutions for each MD subproblem will start the Shaftel-Thompson MD algorithm at an initially good primal feasible solution.

(2). We have stated conditions that limit the search for p. Also, any branch and bound procedure for searching for an optimum solution to a given p, generates a tree which can be used during the solutions procedure for a new value p'. This means that rather than solving several problems from the beginning we may merely extend the tree for each new value of p.

(3). The solutions for $p = 1$ and $p = n$ can be found quickly and give a good initial upper bound on the value of the objective function.

(4). There exist many good local solutions to the problem so that a heuristic search will tend to find a reasonable solution.

REFERENCES

[1] Charnes, A., and M. Kirby, "Modular Design, Generalized Inverses and Convex Programming," Operations Research, Vol. 13, pp. 836-847 (1965).

[2] Duffin, J., E. Peterson, and C. Zener, Geometric Programming, John Wiley and Sons, Inc., 1967.

[3] Evans, D., "Modular Design -- Special Case in Nonlinear Programming," Operations Research, Vol. 11, pp. 637-647 (1963).

[4] Evans, D., "A Note on 'Modular Design' - A Special Case in Non-Linear Programming," Operations Research, Vol. 18 (May-June 1970), pp. 562-564.

[5] Mangasarian, O., Nonlinear Programming, McGraw-Hill Book Co., 1969.

[6] Passy, A., "Modular Design: An Application of Structured Geometric Programming," Operations Research, Vol. 18, pp. 441-453, (1970).

[7] Rutenberg, D. and T. Shaftel, "Product Design: Subassemblies for Multiple Markets," Management Science, forthcoming.

[8] Shaftel, T., "An Integer Approach to Modular Design," Operations Research, Vol. 19, pp. 130-134 (1971).

[9] _____, and G. L. Thompson, "A Simplex-Like Algorithm For the Continuous Modular Design Problem," Management Sciences Research Report No. 248, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa., May, 1971.

[10] Silverman, G., "Primal Decomposition of Mathematical Programs by Resource Allocation: I - Basic Theory and a Direction Finding Procedure," IBM Data Processing Center Report 320-2639 (August 1970).

[11] _____, "Primal Decomposition of Mathematical Programs by Resource Allocation: II - Computational Algorithm with An Application to the Modular Design Problem," IBM Data Processing Center Report 320-2643 (December 1970).

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Graduate School of Industrial Administration Carnegie-Mellon University	Unclassified
	2b. GROUP Not applicable

3. REPORT TITLE

THE CONTINUOUS MULTIPLE MODULAR DESIGN PROBLEM

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Management Sciences Research Report July, 1971

5. AUTHOR(S) *(First name, middle initial, last name)*

Timothy L. Shaftel
G. L. Thompson

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 1971	23	11

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N0014-67-A-0314-0007 b. PROJECT NO. NR 047-048 c. d.	Management Sciences Research Report No. 254 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* WP #5-71-2

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Not applicable	Logistics and Mathematical Statistics Br. Office of Naval Research Washington, D. C. 20360

13. ABSTRACT

In the present paper we extend our recent work on the continuous single module design problem to the multiple module case. It is assumed that there is a fixed cost associated with each additional module used to solve the problem. The Kuhn-Tucker conditions characterize local optima among which is a global optimum. Modules are associated with partitions and a special class, guillotine partitions, are characterized. Branch and bound, partial enumeration and heuristic procedures for finding optimum or good guillotine partitions are discussed and illustrated with examples.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Unclassified

Security Classification

A-31408

AN INTEGER APPROACH TO MODULAR DESIGN

by

Tim Shaftel

May 1971
Carnegie Mellon University
Graduate School of Industrial Administration

ABSTRACT

This paper discusses the modular design problem as set up by David Evans. In this problem, the module is made of several components; there exist x_i of the i th component in each module. For every application, j, y_j modules will be needed to satisfy all component requirements for that application. The problem is to design a module such that the requirements are satisfied at a minimum cost of surplus components. Mathematically, the problem is:

$$\begin{aligned} &\underset{x,y}{\text{Minimize}} \quad \sum_i c_i x_i \quad \sum_j d_j y_j \\ &\text{Subject to} \quad x_i y_j \geq r_{ij} \end{aligned} \qquad \text{for all } i,j$$

x_i, y_j are positive integers and c_i, d_j are constants.

The paper is divided into two parts. The first part introduces the problem and describes the difficulties of rounding the continuous solution in order to achieve an integer solution. Some properties of the constraint set which aid in the integer solution of the problem are also presented. The conclusion of this portion of the paper is that the two procedures outlined by Evans and by Charnes and Kirby for the non-integer case cannot be used to find the integer solution. An integer branch and bound technique is proposed which has been shown to be useful in the solution of small problems and which has potential use in larger problems.

The second part of the paper discusses the use of the objective function to improve the branch and bound procedure. The solution to the continuous problem is also used. A more complete integer solution algorithm is presented as well as a few heuristic rounding techniques.

Part A. An Integer Search Based on the Constraint Set*

*A version of this section has been published under the title "An Integer Approach to Modular Design," in Operations Research, Vol. 19, 1971 (pp. 130-134).

The modular design problem was formulated by David Evans [4]. Evans gave a simplified example and described a solution procedure using nonlinear programming. In 1965, A. Charnes and M. Kirby [3] demonstrated the use of generalized inverses and convex programming in a solution procedure to Evan's problem. Both approaches solve the modular design problem as a continuous valued problem; Evans indicates that, once a non-integer solution was achieved, intelligent rounding would hopefully lead to the integer optimum solution. This note presents a total search procedure, using partial enumeration, as a feasible approach to solving small modular design problems for optimal integer solutions. Applying the total search, it is shown that use of the continuous procedures of Evans, and of Charnes and Kirby for an integer problem will lead to a solution which, in general, is not the optimal integer solution.

The modular design problem is to choose the number of each type of part to place on a single module. Modules are combined to satisfy the needs of each application. The objective is to minimize the total cost of parts used, i.e.,

$$\underset{x,y}{\text{Minimize}} \quad \sum_i c_i x_i \sum_j d_j y_j = \theta$$

$$\text{s.t.} \quad x_i y_j \geq r_{ij}$$

$$x_i, y_j, c_i, d_j, r_{ij} \geq 0$$

$$x_i, y_j \quad \text{Integers}$$

where,

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

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

c_i = cost of part i

d_j = demand for application j

r_{ij} = number of part i units required in application j

x_i = the number of part i on the module (decision variable)

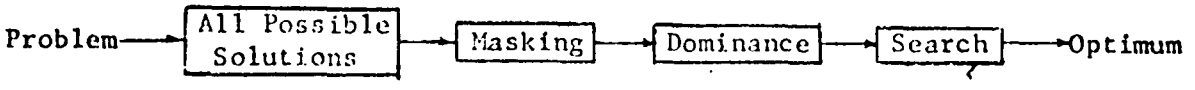
y_j = the number of modules needed in application j (decision variable)

An Integer Search for the Exact Optimum

The example problem posed by Evans is for m=4,n=3:

$$R = \begin{matrix} 15 & 23 & 44 \\ 13 & 13 & 0 \\ 15 & 17 & 35 \\ 34 & 12 & 22 \end{matrix} \qquad c = (1,1,1,1) \qquad d = (1,1,1)$$

The optimum integer solution is x = (23,13,18,34), y = (1,1,2) and $\theta = 352$. To find the exact optimum solution, all possible feasible solutions to the problem were enumerated. There exist 34,408 possible integer solutions to the example problem. Using two bounding techniques, dominance and masking, the problem was solved in about six seconds of computer time.



The total search algorithm simply increments the y values so that all possible n-tuples appear for the y-vector. Once the y-vector is set, the value of the x-vector which minimizes the value of the objective function will be given as:

$$x_i = \underset{j}{\text{Max}} \langle r_{ij}/y_j \rangle \text{ for all } i$$

where $< a >$ indicates the smallest integer value greater than a. It is easily seen that any smaller value of x_i will not satisfy all the constraints of the form $x_i y_j \geq r_{ij}$, while any greater value for x_i will yield a higher valued objective function. It must now be shown that there exists only a finite number of y n-tuples which must be inspected.

Lemma: For any modular design problem with an m by n requirements matrix, the maximum number of solutions which must be searched is $\prod_{j=1}^{m} \underset{i}{\text{Max}} < r_{ij} >$.

Proof: Let $y_k = \underset{i}{\text{Max}} < r_{ik} >$ for some k. Then,

$$x_i = \underset{j}{\text{Max}} \ <r_{i1}/y_1>, <r_{i2}/y_2>, \ldots, <r_{ik}/y_k>, \ldots, <r_{in}/y_n>.$$

$<r_{ik}/y_k> = 1$ for all i and $<r_{ik}/y'_k> = 1$ for $y'_k > y_k$.

If we allow the value of y_k to increase to y'_k while the other y_j values remain constant, the value of the objective function can only increase. For any k, therefore, values greater than $\underset{i}{\text{Max}} <r_{ik}>$ need not be searched; the maximum number of solutions which must be considered will be $\prod_{j=1}^{m} \underset{i}{\text{Max}} < r_{ij} >$.

Reduction of the Search Through Masking and Dominance

Masking will be defined as the property that:

$$x'_i = x_i \quad \text{for all i}$$

while, $y'_j \geq y_j$ for all j and, $y'_k > y_k$ for at least one k.

This situation will cause the objective function to increase as y_k increases. In the Evans example, for instance, the solutions y = (15,13,12), y= (15,13,13), and y = (15,13,14) all have the same minimal x-vector of x = (4,1,3,3). Since the objective function can only increase as y_3 increases, the last two y-vectors are masked out by the first. Dominance is similar to masking; it will be defined as the

property that:

$< r_{ik}/y_k > \leq <r_{ij}/y_j>$ for all i and any $j \neq k$.

In the case where $< r_{ik}/y_k >$ is dominated, any value larger than y_k will not affect the choice of x (i.e., $x_i = \underset{j \neq k}{\text{Max}} < r_{ij}/y_j >$).

Again taking Evans' example and using y = (1,1,3), we have x = (23, 13,17,34). For each i, the value r_{i3}/y_3 is completely dominated by the values of $< r_{i1}/y_1 >$ and $< r_{i2}/y_2 >$; the y-vectors (1,1,3), (1,1,4), (1,1,5), etc. will have the exact same x-vector of (23,13,17,34) and a higher valued objective function as y_3 increases from 3. Dominance then is a specific type of masking whereby, after y_k exceeds a certain value, all higher values are masked if the remaining y_j values are held constant. The use of dominance and masking in the example problem reduced the search to about 850 possible solutions. Dominance and masking are types of bounding techniques tailored to this problem to eliminate from consideration groups of solutions which have higher costs than those already found. This procedure is similar to the partial or implicit enumeration technique presented by Egon Balas ([1] and [2]) where certain branches of the solution tree are curtailed or eliminated so that all elements of the solution set are implicitly examined.

Difficulties With Non-integer Solution Procedures Applied to Integer Modular Design Problems

Both Evans' and Charnes and Kirby's procedures use the transformation $s_i = c_i x_i$, $t_j = d_j y_j$ and $r'_{ij} = c_i d_j r_{ij}$ in order to simplify the problem structure. When the problem has integer constraints, however, this transformation cannot be made. The optimum integer solution to the original problem will not be the same as the optimum integer solution to the transformed problem. The difficulty arises from the fact that, in order

to attain the minimum, we have $x_i = \underset{j}{\text{Max}} < r_{ij}/y_j >$ which is, in gneral, a different value than x_i calculated from the equation $s_i = \underset{j}{\text{Max}} < r'_{ij}/t_j >$ which becomes $c_i x_i = \underset{j}{\text{Max}} < c_i r_{ij}/y_j >$. For example, let $r_{ij}/y_j = .196$ be the maximum ratio over j; x_i should equal one. Inspection of the transformed problem,however, shows that the calculated value of x_i would be dependent on the value of c_i, $c_i x_i = < .196\ c_i >$ or $x_i = 1/c_i < .196\ c_i >$. For instance, setting $c_i = .10$, the result would be $x_i = 10\ x < .0196 > = 10$ which is clearly incorrect. As a further example, form two problems P1 and P2 using the following data:

P1: $R = \begin{matrix} 15 & 23 & 44 \\ 13 & 13 & 0 \\ 15 & 17 & 35 \\ 34 & 12 & 22 \end{matrix}$ $c = (1,2,3,4)$ $d = (1,2,3)$

P2: $R = \begin{matrix} 15 & 46 & 132 \\ 26 & 52 & 0 \\ 45 & 102 & 315 \\ 136 & 96 & 264 \end{matrix}$ $c = (1,1,1,1)$ $d = (1,1,1)$

The transformed problem for both P1 and P2 will be exactly the same, the corresponding non-integer optimum solutions, therefore, being equal. Using the exact search algorithm, however, the optimum integer solution to the first problem is x = (23,13,18,12), y = (3,1,2), θ = 1661, while the second problem yields an optimum integer solution of x = (4,4,9,8), y = (17,13,35), θ = 1625. It may be concluded from this example that rounding the non-integer optimum solution in a attempt to locate the optimum integer solution cannot be accomplished as Evans had felt.

Conclusions

The approaches to the continuous modular design problem proposed by D. Evans and by A. Charnes and M. Kirby cannot locate the exact optimum solution for the case where the values are constrained to be

integers. A computer approach which first reduces the number of feasible solutions and then searches these solutions for the optimum can find the exact optimum to a small problem. The 34,408 solution example problem was solved in six seconds on the Univac 1108; a second problem with 4,359,480 solutions was solved in eleven seconds.

It is estimated that the present computer program can be used to solve 4 by 10 variable problems with requirements of up to 500 units. Additional modifications using other branch and bound techniques may make this method reasonable for larger problem solutions. One last observation, which will add to the possibilities of using this exact search, is that the x and y values may be constrained not to exceed certain amounts. The exact search is particularly well suited for this type of bounding which would make it feasible for problems larger than four by ten.

Part B. An Integer Search Using the Continuous Solution

The solution to the continuous modular design problem is a ratio in both x and y. That is, given a solution $\bar{x}$, $\bar{y}$, then $\bar{x}/t$ and $\bar{y}\cdot t$, $t > 0$, is also a solution and yields the same optimum [4]. In a few specific cases the solution to the integer problem follows directly from the solution to the continuous problem. This case arises whenever the lowest common denominator of the fraction y_j, for all j (using $\sum_j y_j = 1$) is a common divisor of x_i, for all i. For instance, the problem:

$$R = \begin{matrix} 1 & 2 & 3 \\ 1 & 9 & 6 \\ 2 & 7 & 16 \end{matrix} \qquad c = d = (1,1,1)$$

has a continuous solution of y = (1/8,3/8,4/8), x = (8,24,32), θ = 64. The integer solution (using t = 8) of y= (1,3,4), x = (1,3,4), θ= 64, is optimal. It is clear that we are searching integer solutions whose y_j ratios are, in some sense, "close" to the optimum ratio of the continuous solution.

A Lower Bound of Partial Integer Solutions

The objective function of the modular design problem is $\theta = \sum_i c_i x_i \sum_j d_j y_j$. We are going to assume that certain of the y_j values have been set to integer values (for convenience, we shall assume that y_j , j = 1,2,...,s-1 are fixed). We are now interested in fixing the value of y_s and determining a lower bound on the value of the objective function. For ease of notation, we denote the ratio y_s/y_j as v_{sj} for j = 1,2,...,s. Clearly $v_{ss} = 1$.

Now using the constraint set, we have:

(1) $x_i y_j \geq r_{ij}$ j = 1,...,s-1 and all i

<=> (2) $x_i y_s \geq v_{sj} r_{ij}$ $\quad j = 1,\ldots,s-1$ and all i

also, (3) $x_i y_s \geq r_{is}$ $\quad$ for all i

<=> (4) $x_i y_j \geq r_{is}/v_{sj}$ $\quad j = 1,\ldots,s-1$ and all i

We have, therefore, using (1) and (4):

$$x_i y_j \geq r_{ij}$$

and $$x_i y_j \geq r_{is}/v_{sj} \qquad j = 1,\ldots,s-1 \text{ and all } i$$

or (5) $x_i y_j \geq \max (r_{ij}, r_{is}/v_{sj})$ $j = 1,\ldots,s-1$ and all i

and, using (2) and (3):

$$x_i y_s \geq v_{sj} r_{ij} \quad j = 1,\ldots s \text{ and all } i$$

or (6) $x_i y_s \geq \max\limits_{j=1 \text{ to } s} (v_{sj} r_{ij})$ for all i

Equations (5) and (6) are tight at the optimum since we are minimizing the value of the objective function.

Let θ'_s be the partial value of the objective function obtained by setting y_j, $j = 1,\ldots,s$ at integer values. Then,

$$\theta'_s = \sum_i \left[\sum_{j=1}^{s-1} c_i d_j \max (r_{ij}/v_{sj},\ r_{ij}) + c_i d_s \max_{j=1 \text{ to } s} (v_{sj} r_{ij}) \right]$$

Let $\theta_s = \theta'_s + \sum_i \sum_{j=s+1}^{n} c_1 d_j x_i y_j$, where the second term is the continuous solution for the subproblem using $j = s+1,\ldots,n$. Then it is easy to see that θ_s is a lower bound on any integer solution which can be achieved from setting the remaining $s+1,\ldots,n$ values of y to integers.

Upper and Lower Bounds

Substituting $v_{sj} = y_s/y_j$, $j = 1,\ldots,s$ into the expression for the partial objective function yields:

$$\theta'_s = \sum_i \left[\sum_{j=1}^{s-1} c_i d_j \max(r_{ij} y_j/y_s,\ r_{ij}) + c_i d_s \max_{j=1 \text{ to } s} (y_s r_{ij}/y_j) \right]$$

Lemma: θ_s is convex in $y_s \geq 0$.

Proof: The continuous term in θ_s does not contain y_s. We must show, therefore, that θ'_s is convex.

The second term of θ_s' is either linear or constant in y_s and is therefore convex. The first term is the maximum of two convex functions of y_s and therefore convex [8, p. 32].

θ_s is infinite at $y_s = 0$ and at $y_s = \infty$. This fact, along with the Lemma implies that, if we have an upper bound on the value of the objective function, θ_n, then we can develop upper and lower bounds on the possible values of y_s (see Figure I). Letting y_s^b and y_s^u represent

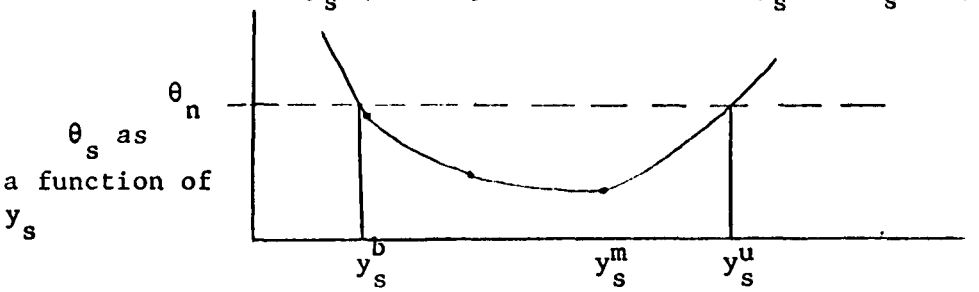


Figure I: Upper and Lower Bounds on y_s

these bounds we have that:

(6) $y_s^b = [y_s \mid \theta_s(y_s) \leq \theta_n \text{ and } \theta_s(y_s-1) \geq \theta_n]$

(7) $y_s^u = [y_s \mid \theta_s(y_s) \leq \theta_n \text{ and } \theta_s(y_s+1) \geq \theta_n]$

The function θ_s is ditonic in y_s (it is the negative of a unimodal function). We can, therefore, use the efficient Fibonacci search to locate the minimum θ_s associated with an integer value of y_s. (Let y_s^m be this value). Once y_s^m is found we need only search $y_s^m + p$ and $y_s^m - q$, p and q integers until equations (6) and (7) are satisfied; no other value of y_s are possibly optimum for this partial solution j = 1,...,s-1.

As an example take the following problem:

$$R = \begin{matrix} 1 & 3 & 3 & 4 \\ 2 & 9 & 8 & 10 \\ 2 & 10 & 16 & 16 \\ 4 & 12 & 18 & 25 \end{matrix} \qquad \begin{matrix} c = (1,1,1,1) \\ \\ d = (1,1,1,1) \end{matrix}$$

with $y_1 = 1$ and $y_2 = 3$.

Then $y_3 = v_{31}y_1$ and $y_3 = v_{32}y_2$

Equations (1):

$$x_1y_1 \geq 1 \qquad x_1y_2 \geq 3$$
$$x_2y_1 \geq 2 \qquad x_2y_2 \geq 9$$
$$x_3y_1 \geq 2 \qquad x_3y_2 \geq 10$$
$$x_4y_1 \geq 4 \qquad x_4y_2 \geq 12$$

Equations (2):

$$x_1y_3 \geq 1v_{31} \qquad x_1y_3 \geq 3v_{32}$$
$$x_2y_3 \geq 2v_{31} \qquad x_2y_3 \geq 9\ v_{32}$$
$$x_3y_3 \geq 2v_{31} \qquad x_3y_3 \geq 10v_{32}$$
$$x_4y_3 \geq 4v_{31} \qquad x_4y_3 \geq 12\ v_{32}$$

Equations (3):

$$x_1y_3 \geq 3$$
$$x_2y_3 \geq 8$$
$$x_3y_3 \geq 16$$
$$x_4y_3 \geq 18$$

Equations (4):

$$x_1y \geq 3/v_{31} \qquad x_1y_2 \geq 3/v_{32}$$
$$x_2y_1 \geq 8/v_{31} \qquad x_2y_2 \geq 8/v_{32}$$
$$x_3y_1 \geq 16/v_{31} \qquad x_3y_2 \geq 16/v_{32}$$
$$x_4y_1 \geq 18/v_{31} \qquad x_4y_2 \geq 18/v_{32}$$

Now $v_{31} = 3v_{32}$, so that we have, for instance:

$$x_2y_2 = \max\ (\ 9,\ 8/v_{32}\ )$$

$$x_2y_1 = \max\ (2,\ 8/3v_{32}\ )$$

and $x_2y_3 = \max\ (8,\ 6v_{32}, 9v_{32}) = \max\ (8,\ 9v_{32})$

For the entire matrix we have (j = 1,2,3):

$$x_iy_j = \text{Max} \begin{matrix} 1,1/v_{32} & 3,3/v_{32} & 3,3v_{32} \\ 2,8/3v_{32} & 9,8/v_{32} & 8,9v_{32} \\ 2,16/3v_{32} & 10,16/v_{32} & 16,10v_{32} \\ 4,6/v_{32} & 12,18/v_{32} & 18,12v_{32} \end{matrix}$$

At $y_3 = 1$, $v_{32} = 1/3$, $\theta_3' = 3+8+16+18+9+24+48+54+3+8+16+18 = 225$.

The solution to the continuous problem for j =4 is simple in this case ($y_4 = 1$, x = (4,10,16,25)). We have, therefore, $\theta_3 = 225 + 55 = 280$. Assume that we know one integer solution exists with $\theta_n = 198$ [i.e.,

$y = (1,2,3,3)$, $x = (2,5,6,9)$]. y_3^b and y_3^u can be determined from the following table:

y_3	θ_3'	θ_3	
0	infinite	infinite	
1	225	280	y_3^b
2	138	193	
3	133	188	
4	104	159	
5	108	163	
6	120	175	
7	132	187	
			y_3^u
8	144	199	
9	156	211	

We also know that the best integer point is $y_3 = 4$. As an aside, using the dominance property discussed in part A, the partial solutions $x = (1,3,6)$ and $x = (1,3,7)$ may be eliminated from consideration.

The Algorithm

The solution algorithm uses the properties outlined in both parts A and B in a partial enumeration scheme similar to [1]. The algorithm requires an efficient solution to the continuous modular design problem. Shaftel and Thompson [7], predict solutions to extremely large problems with little computer time. In the following algorithm we shall assume that all y_j, $j = 1,\ldots,s-1$, have been set to integer values.

(0) Initialization Step. Find an initial integer solution and its objective function, θ_n. Techniques for heuristically developing a good solution to this problem will be discussed in the next section. Set the initial value of y_1 at one. Set $s = 2$ and go to (1).

(1) Forward Step. Using a Fibonacci Search find the optimum

integer value of y_s; set y_s to this value. Go to (2).

(2) Test Step. Find

a. If $\theta_s \geq \theta_n$, go to (3).

b. If $\theta_s < \theta_n$, and $s < n$, set $s = s + 1$ and go to (1).

c. If $\theta_s < \theta_n$, and $s = n$, go to (4).

(3) Backtrack Step. Free y_s from its integer constraint.

a. If $s = 2$, let $y_1 = y_1 + 1$. If y_1 is now greater than $\max_i r_{i1}$, stop--the solution association with θ_n is optimum otherwise go to (1).

b. If $s \neq 2$, change y_{s-1} to some other possible value such that $y^b_{s-1} < y_{s-1} < y^u_s - 1$. Do this by increasing or decreasing y_{s-1} by a value of one and check to see if this possibility is masked or dominated by some other value. Eliminate, or course, any previously checked value of y_{s-1}. Note that the values of y^b_{s-1} and y^u_{s-1} need not be calculated in advance. We will know when these bounds have been exceeded by the value of $\theta_s(y_s)$. If there is no value between the bounds which has not been eliminated, let $s = s-1$ and go to (3). Otherwise let $s = s-1$ and go to (2).

(4) Improved Solution Step. Let $\theta_n = \theta_s$. The new optimum solution is the one which corresponds to this value of the objective function.

Lemma: Use of the above algorithm will locate the minimum integer solution.

Proof: For every partial integer solution for y_j $j = 1,2,\ldots,s-1$, all values of y_k $k = s,s+1,\ldots,n$ are either enumerated or eliminated from consideration by bounding, masking or dominance. Since y is incremented to

all its possible values the algorithm attains the minimum.

Certain portions of the basic algorithm may be modified with possible computational improvements. Although the results of these modifications must be tested, a few of these changes will be discussed at this time.

1) The speed of the algorithm could depend to a great extend on the column order used during computation. One ordering with this in mind might be to put columns with small $\max_i r_{ij}$ in the beginning. The advantage to this procedure arises from the fact that y_j's which are set early in the procedure are likely to take more of their possible values (1 to $\max_i r_{ij}$). Another ordering would be to put columns which are likely to create large costs early in the procedure, thus improving the bounds on later y_j's. Naturally, some trade off between the two orderings suggested would be made.

2) A second modification could be made on the lower bound θ_s to an integer solution. This could be accomplished by forcing y's which have already been set at integers to assume the same ratios for the continuous problem which included all the columns. We could then solve for the minimum of the continuous problem, with certain cells forced to remain tight, as a lower bound to the overall problem. The disadvantage here would be that, at every step, we would have to solve a new continuous problem. The gain, of course, would be the possibility of a better lower bound.

3) The final modification which will be discussed is the possibility of setting more than one integer value at a time. To be sure this would require a more complex search than the relatively simple Fibonacci search for setting one variable. The advantage of this approach would be the efficiencies attained by setting two or more integer values simultaneously and the possibility of finding better solutions early in the procedure.

Heuristic Schemes for a Good Integer Solution

We will discuss three approaches for attaining a quick heuristic integer solution. All three approaches use the continuous solution to achieve an integer solution which is close to the optimum ratio.

The first approach is based on the fact that the continuous solution has $\sum_j y_j = t$, where t is an integer value (t = 1 is usually chosen for the solution to the continuous problem). By choosing integer values of t and rounding y we can develop an integer solution. For example, assume y = (.19,.34,.57) for the optimum continuous solution. Then, letting t = 10 yields y = (1.9,3.4,5.7) and y = (2,3,6) as an integer solution. Several arbitrary choices for t would probably be sufficient for a good initial solution. A good selection of t would be to choose values such that x and y will be about the same order of magnitude. In this way, rounding small numbers will, as much as possible, be prevented.

The second approach would be to choose one y_j and to set its value to an integer. The continuous solution would then be used to determine the other y_k, $k \neq j$ values to be rounded to integers. Since y_j is less than $\max_i r_{ij}$ for any j, this heuristic could set y_j to any value between one and $\max_i r_{ij}$.

The third approach is similar to the second except that the y_k, $k \neq j$ values are set one at a time; as each value is set a new continuous solution is found which is then used to set the next value. Examples of the first two techniques will be shown. We will solve the following example problem:

$$R = \begin{matrix} 1 & 5 & 9 & 8 \\ 1 & 7 & 12 & 4 \\ 0 & 3 & 8 & 10 \\ 1 & 4 & 6 & 12 \end{matrix} \qquad c = d = (1,1,1,1)$$

The continuous optimum solution to this problem is y = (.039, .198,.339,.424): x = (25.3,35.4,23.6,28.3), θ = 112.6. Using t = 5 to 15 yields the following table:

t	ty_1	ty_2	ty_3	ty_4	y_1	y_2	y_3	y_4	θ
5	<1	<1	1.69	2.11	1	1	2	2	138
6	.	.	.	2.45	1	1	2	3	140
7	.	.	.	.	1	1	2	3	-
8	.	1.58	2.71	.	1	2	3	3	135
9	.	.	.	3.81	1	2	3	4	130
10	.	.	.	.	1	2	3	4	-
11	.	3.73		4.66	1	2	4	5	144
12	.	.	.	.	1	2	4	5	-
13	.	2.57	.	5.51	1	3	4	6	140
14	.	.	4.77	.	1	3	5	6	145
15	.	.	.	.	1	3	5	6	-

Note that the ratio of the best integer solution found to the optimum continuous solution is 130/112.6 = 1.15

The second technique would set $y_1 = 1$, then

$$y_2 = .198/.039y_1 = 5.0$$
$$y_3 = .339/.039y_1 = 8.6$$
$$y_4 = .424/.039y_1 = 10.7$$

This means that y = (1,5,9,11), x = (1,2,1,2), 0 =156. Rounding y_4 to 10 would yield y = (1,5,9,10), x = (1,2,1,2), 0 = 150.

No other value of y_1 need be considered. It should be pointed out that if the first technique were used on the x-vector of the continuous problem with (t =1/25) the solution would be x = (1,1,1,1), y = (1,7,12,12), θ = 128 which is the optimum integer solution for this problem.

Conclusions

By use of the continuous solution to the modular design problem an integer search algorithm can be improved. The algorithm presented in this paper has several advantages.

1. The continuous solution to any subproblem made of columns s + 1 to n remains the same throughout the algorithm. This means that only n continuous subproblems must be solved. Since the subproblems get smaller and smaller, since parametric procedures exits, and since fast continuous procedures are available, this portion of the algorithm is very efficient.

2. The algorithm procedure concentrates only on y (where y is the variable of smallest dimension). As the dimension of x increases, better bounds on y can be achieved so that computer time should not increase any more than, at worst, linearly, in the dimension of x.

3. The upper and lower bounds on y seem to be fairly tight. Masking and dominance also seem to be efficient in reducing the search space. Though the search space will remain large, it will be nowhere near the possible value of $\prod_{j=1}^{n} \max_i r_{ij}$.

4. The algoritm can be designed to look at integer solutions close to the optimum continuous solutions quickly. This means that stopping the algorithm before the enumeration is complete would yield a good solution. The continuous solution provides us with a good lower bound on the optimum integer solution; this lower bound can be used to evaluate any integer solution.

5. The nature of the problem is such that the requirements matrix is normally long and thin (i.e., we have many part but few applications). This means that, although the search over y probably increases fairly fast in the dimension of y, we must search over a relatively small number of dimensions.

6. Finally, the algorithm is adaptable. The addition of costs which are a function of y (such as handling costs for placing more modules in applications) would reduce the search since large values of y would become intolerable. Constraints on the magnitude of both x and y vectors would also reduce the possible search space.

The algorithm is also adaptable to a heuristic procedure for dealing with more than one module. In this instance, we could determine the continuous solution of each module and use the algorithm on these modules, to determine the integer solutions, one module at a time.

REFERENCES

[1] Balas, Egon, "An Additive Algorithm for Solving Linear Programs with Zero-One Variables," Operations Research, Vol.13, pp. 517-546 (1965).

[2] _____, "Discrete Programming by the Filter Method," Operations Research, Vol. 15, pp. 915-957 (1967).

[3] Charnes, A. and M. Kirby, "Modular Design, Generalized Inverses and Convex Programming," Operations Research, Vol. 13, pp. 836-847 (1965).

[4] Evans, David H., "Modular Design - Special Case in Nonlinear Programming," Operations Research, Vol 11, pp.637-647 (1963).

[5] Passy, U., "Modular Design: An Application of Structured Geometric Programming," Operations Research, Vol. 18, pp. 441-453 (1970).

[6] Rutenberg,D., and T. Shaftel, "Product Design: Subassembles for Multiple Markets," Management Science, to appear.

[7] Shaftel, T. and G. Thompson, "A Simplex-Like Algorith for the Modular Design Problem," Management Science Research Report No. 248, Graduate School of Industrial Administration, Carnegie Mellon University, (May 1971).

[8] Zangwill, W.,Nonlinear Programming A Unified Approach, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1969.