The Theory of Linear Economic Models

DAVID GALE

Associate Professor of Mathematics, Brown University and
Consultant to Mathematics Division of The RAND Corporation

McGRAW-HILL BOOK COMPANY

New York *Toronto* *London* 1960

THE THEORY OF LINEAR ECONOMIC MODELS

 Library of Congress Catalog Card Number: 59-15461

9 10 11 – M P – 9 8

22728

Preface

This book is written at a time of revived activity in the field of applied mathematics. "Revived" is perhaps the wrong word to use in this connection, for the characteristic feature of the new applied mathematics is not an intensification of work on old problems but rather an attempt to extend the application of mathematical reasoning to entirely new kinds of situations. Information theory, cybernetics, game theory, theory of automata are but a few of the new disciplines. Naturally, much of the work in these subjects is of a tentative and experimental nature. On the other hand, there have been certain developments which after a decade's experience seem to be of permanent usefulness. One such is probably information theory. Another is linear programming and the related linear models. Being convinced that this latter subject is "here to stay," I felt it was appropriate to try preparing a suitable text. This book is the result.

Before asking the reader to plunge into the subject of linear models I shall, in accordance with a sensible custom, attempt in the few pages which follow to give some idea of what this subject is. An ideal preface is one which tells the reader in a few words exactly what the rest of the book contains and thus saves him the trouble of reading it. I regret that the writing of such a preface in the present case is beyond my powers of exposition. The best I can do is to describe in a general way the sort of problems we shall be concerned with, the approach we shall take to these problems, and the manner in which the relevant material will be organized.

The Subject Matter. The term "economic model" is admittedly a vague one, but for our purposes we may think of such a model as an abstraction and simplification of some typical economic situation. As an example, the first model we shall take up is that of linear programming, which in its abstract formulation is a certain kind of mathematical maximum or minimum problem. The importance of this model derives from the fact that many actual economic situations lead to precisely this problem after the appropriate simplifying assumptions have been made. Later we take up the two-person game model. This is again formulated in a purely abstract manner, but the significance of the model for us comes from the fact that it is designed to reflect the essential features of certain games of strategy, and thus indirectly certain aspects of economic competition. Other models to be treated concern patterns of exchange between countries or industries, alternative schemes of production, certain economic equilibrium situations, and so on. In each case the models will be introduced by first describing the economic situation, next stating what simplifications are to be made, and then giving the purely abstract formulation.

Having arrived at this abstractly formulated model, what do we intend doing with it? By way of answer let us first state clearly some of the things we *do not* intend doing. A very important question in relation to any model is that of applicability. Does the model really give a reasonable approximation to the situation which gave rise to it? Is it to be relied on in making decisions and predictions? To what extent have predictions based on the model been borne out experimentally? Such questions belong to pure economics and will not be touched on here. Indeed, the models we have chosen to discuss vary widely as regards applicability. At one extreme we have linear programming, which is already being used quite extensively in industrial planning. At the other we have topics like game theory and some of the equilibrium models, which are in no sense ready for practical application in their present stage of development.

But if applicability is not the criterion for selection how then have we decided which topics to discuss? The answer is this: We have tried to select those models which best illustrate the manner in which mathematical reasoning can be used to obtain information about idealized economic situations. In some instances we have had to make rather drastic simplifications. The resulting lack of realism is unfortunate but is to be expected in early attempts at understanding complex situations.

Having formulated our models, the rest of the task consists in analyzing them, that is, of deducing in a rigorous fashion the consequences of the assumptions which have been made. The procedure is quite analogous to deducing theorems from the axioms of, say, plane geometry. As in the case of geometry, some of the results we shall obtain could hardly have been guessed in advance. It is this fact which encourages one to believe that mathematical analysis may help to bring about new and significant advances in the understanding of economic phenomena.

We have restricted our presentation to the study of *linear* models, that is, roughly speaking, models in which the mathematical relations have the form of equations or inequalities of degree one. This restriction is due simply to limitations of space and time. An equal number of pages could have been devoted to nonlinear models. This would, however, have involved developing a great deal of additional mathematical machinery, and for this reason we chose to remain within the linear framework. A further justification for this decision was the fact that most of the nonlinear results make use of the linear theory. Much of this book may thus be regarded as foundation material for work on more advanced levels.

It might be thought from what has been said so far that we have gathered together a miscellaneous collection of problems whose only common features are an economic flavor and the occurrence of linear relations. Fortunately, this is not the case, for although there is considerable variety in the models to be studied, the mathematics involved will exhibit a noteworthy degree of unity. Most of our analysis will use the mathematical material developed in Chap. 2 on Real Linear Algebra or, in more everyday language, the theory of linear equations and inequalities in real numbers. The feature of this theory which plays the unifying role in most of the applications is the fundamental notion of *duality*. We shall not even attempt to define this term here but remark that it is the recurrent theme which ties together the various parts of the book into what may legitimately be called a theory.

The Approach. We have already remarked that this book is intended as a text. We hesitate to use the words "advanced text," for this suggests that preliminary familiarity with the subject matter is assumed, which is not the case. The book is advanced in the sense that it attempts to bring the reader to the frontiers of the subject, enabling him to understand and possibly contribute to current research in the field. In other words, we are trying primarily to fill the needs of the would-be specialist, be he mathematician, economist, business student, or engineer. But while our main objective is the training of experts, we have tried to arrange matters so that the book will also be useful to readers who wish to go into the subject less intensively. The less technical parts of the book, in particular Chapter 1 on linear programming and most of Chapter 6 on game theory, are designed to be usable in courses on these subjects on the level of an advanced undergraduate course in economics or engineering.

Concerning the use of the book as a basic text for a course, it should be explained that the book is itself based on a set of notes from a course given to a group of graduate students in pure and applied mathematics, and the treatment should be suitable for students at this level. We suspect the average graduate student in economics would have some difficulty in going through the book on his own, for we emphasize that this is a text not in economics but in applied mathematics. Nevertheless, the theorems we prove are about economics, are used by economists, and in many cases were first discovered by economists.

Concerning the use of this book by economists, a further word of caution is in order. It has been brought to my attention by Professor Dorfman that certain words and expressions mean quite different things to economists on the one hand and mathematicians on the other. It was both startling and illuminating to me to realize that the very first words of my title "The *Theory* of" belong to this category. By way of illustration, a mathematician or natural scientist on reading one of the important *theory of* books of economics, say Hicks or Keynes, might well remark "very interesting, but where is the theory?" The remark would imply no disparagement of these works but would simply point up a confusion of language, for the natural scientist expects a theory to consist of a large body of results derived from a small set of assumptions. What he has read consists instead of a careful formulation and detailed justification of a particular set of assumptions, with rather less formal deduction of implications than he would find in a theoretical treatise in the natural sciences. Analogously, an economist reading the present volume will undoubtedly feel that it has been misnamed in that most of the "theory" has been left out, and he will correctly point out that the book is teeming with economic assumptions for which little or no justification is given. We reply that the word "theory" is to be understood here as it is used in the natural rather than the behavioral sciences and is therefore not directly concerned with the justification of assumptions. We stress this point in order not to mislead the reader concerning our intentions.

It is our hope that our presentation of results will be useful to the economics student with exceptional aptitude for the mathematical approach. It should also be useful in the hands of a teacher of mathematical economics who can modify the exposition to suit the needs of his students, skimming over portions which present purely technical difficulties, elaborating on other parts in which our treatment has not been sufficiently detailed. As such this book might usefully supplement one of the texts in economics which covers the same material, such as "Linear Programming and Economic Analysis" by Dorfman, Samuelson, and Solow or "Mathematical Economics" by R. G. D. Allen.

Finally, we hope the book will be used as a reference for workers in the field of linear models who will find here a mathematically unified treatment of many important results which were previously available only in scattered sources in the economic and mathematical literature.

We come next to the question of mathematical prerequisites. It is customary to remark at this point that the only requirement for an understanding of what is to follow is a knowledge of elementary calculus. In the present case even this requirement may be waived, for calculus is never used. Our principal tool is matrix algebra, but no previous knowledge is required here either, as all necessary facts are developed in the text. What is required is the ability to follow a moderately involved mathematical argument, an ability which generally comes only with a fair amount of experience and is often characterized by the illusive phrase "mathematical maturity." Some of the proofs we shall present are quite difficult. Even the proof of the "theorem of the separating hyperplane," which is the key mathematical result of the book, is not entirely straightforward. There is no way around this difficulty, for most of the results we wish to present are not mathematical trivialities, and one cannot make things easy without omitting proofs altogether, which would defeat our main purpose. We shall, of course, use all the available devices to help the reader's understanding such as geometric pictures, plausibility arguments, and numerical examples.

We may summarize what has been said in the foregoing paragraphs by remarking that a course based on this book would occupy a position somewhat analogous to a course in mathematical statistics. Such courses are generally given in a mathematics department but are available to students in other fields with the necessary mathematical qualifications.

The Organization. How to Use the Book. We envision four possible courses which could be based on this book.

1. A full-year course covering the entire nine chapters. It would not be necessary to take them up in order, as will be seen from the diagram on page ix.

2. A one-semester course on linear programming. This would cover the first five chapters of the book.

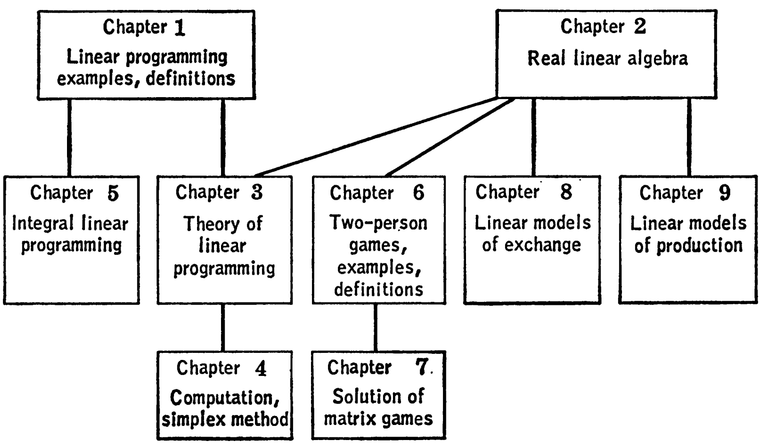
3. A one-semester course in linear programming and game theory. This would consist of Chaps. 1, 2, 3, 6, and 7, omitting Sec. 2 of Chap. 7.

4. A one-semester course in linear economic models. This would cover Chaps. 1, 2, 3, 8, and 9.

The schematic diagram on page ix shows how the various chapters depend on each other.

As the figure shows, Chap. 2 on Real Linear Algebra is necessary for all later chapters. However, the second half of the chapter, from Sec. 5 on, is used only occasionally in subsequent chapters. The instructor may wish, therefore, to take up only the first four sections of this chapter, which are sufficient for all the applications in Chaps. 3, 4, 6, and 8.

From a logical point of view it would have been most natural to begin with Chap. 2, in which the mathematical machinery is developed. This procedure would have the disadvantage, however, of requiring the reader to absorb a considerable amount of abstract material without knowing what it was to be used for. For this reason it seemed preferable to start with the applications, in this case linear programming, and state the main theoretical results without proof in order to motivate further study in the algebraic foundations. Chapter 1 is therefore devoted to describing the linear programming problem first by means of a set of illustrative examples, then by a formal definition. The discussion of the next section leads up to the state-



ment (but not the proof) of the fundamental duality theorem, which is then illustrated in specific cases. Assuming the duality theorem we then prove the important "equilibrium theorem" and give applications. The chapter, like all the others, ends in a short set of bibliographical references and a somewhat longer set of exercises of varying degrees of difficulty.

The first sections of Chap. 2 are devoted to introducing vectors and matrices and developing the classical theory of linear equations in a rapid but complete and self-contained manner. The mathematical heart of the chapter, and, in fact, of the book, is in Secs. 3

and 4, in which we develop the not so classical theory of real linear equations and linear inequalities. The latter half of the chapter is devoted to a more detailed and somewhat geometric analysis of the solutions of inequalities.

The reader who is acquainted with linear algebra may be struck by the fact that certain popular topics in this subject are conspicuous by their absence, among them the theory of determinants and of characteristic roots or eigen-values. The reason for this omission is simply that we know of no cases in which these particular algebraic objects are useful in drawing conclusions about economic models, and therefore there is no reason why the reader should spend time trying to master these somewhat intricate topics.[1]

In Chap. 3 we return to linear programming problems, which are now defined in complete generality. Using the algebraic apparatus developed in Chap. 2 it is possible to give a complete treatment of the duality and equilibrium theorems as well as the important result on basic solutions. The last part of the chapter is concerned with a most important economic application of linear programming theory, namely, the solution of the problem of optimal resource allocation by the method of price equilibrium under free competition.

Chapter 4 is devoted primarily to an exposition of the simplex method of Dantzig and its application not only to linear programming but also to such general problems as solving systems of inequalities and finding nonnegative solutions of linear equations. Our approach has been to show that the simplex method may be looked upon as an extension of the ordinary "high-school method of elimination" for solving sets of simultaneous linear equations. In vector language the

[1] In view of the rather frequent occurrence in the economic literature of results involving determinants and eigen-values, this statement perhaps calls for some amplification. An example will perhaps illustrate the point. It is a true theorem that a Leontief model is capable of producing a positive bill of goods if and only if the principal minors of the production matrix are all positive. This fact, however, gives us no new economic insight into the properties of Leontief models because there is no economic interpretation to be attached to these principal minors. Contrast this result with the theorem that if a Leontief model can produce one positive bill of goods it can produce any positive bill of goods. The latter statement is a useful and interesting result about the model itself, since both the hypothesis and the conclusion have an obvious economic meaning.

elimination of a variable becomes the replacement of a vector in a basis, and it is this "replacement operation" which becomes the basic computational unit in our presentation. The final section of the chapter presents the generalized simplex method of Dantzig, Orden, and Wolfe for resolving the problem of degeneracy.

Chapter 5 is devoted to the very important class of linear programs, including transportation problems, which always have integral solutions if the initial data are integral. As indicated by our schematic diagram, the material of this chapter is essentially independent of the previous theory. We begin by presenting the network-flow theory of Ford and Fulkerson which, together with the method of Kuhn for the optimal-assignment problem, provides us with a complete and elegant theory for a wide class of integral problems. The relationship of this theory to the classical notion of price equilibrium is given in Sec. 6. The Hitchcock transportation problem is treated in detail as well as various other applications. Again in this chapter it is the duality concept which does the work.

In Chap. 6 we introduce two-person zero-sum games by a sequence of examples which lead first to the statement and then the proof of von Neumann's minimax theorem. The proof is that of Gale, Kuhn, and Tucker using the symmetrization of a game of von Neumann.

The "equivalence" of linear programming and matrix games is the first topic of Chap. 7, and it is shown that the minimax theorem can be derived as a special case of the fundamental duality theorem of linear programming. A short section is devoted to solving games by the simplex method. Several sections are then devoted to a detailed analysis of the structure of the sets of optimal strategies of a matrix game. The final sections are devoted to a description of the method of fictitious play of Brown and to Robinson's proof that the method converges.

Chapter 8 is concerned first with a linear exchange model, equivalent versions of which seem to have been discovered independently by Frisch, Remak, and Bray. A complete analysis is given of the equilibria of such models. A dynamic theory of linear trade models is then treated along the lines of some work of Solow. The theorems here are exactly the same as those which occur in the theory of Markov chains in probability theory. The final sections of the chapter treat a particular model of price equilibrium.

Among the topics treated in the final chapter are Leontief models, including the Samuelson-Koopmans-Arrow substitutability theorem, the work of Koopmans on the relation between efficiency and profit maximization, and von Neumann's expanding linear model.

Terminology, Notation, Bibliography. We shall, of course, define all technical terms and symbols as they are introduced. For the most part we have adhered to standard terminology and notations when such things existed. On the other hand, we have exercised the mathematical equivalent of poetic license to institute an occasional "improvement," mostly in the interests of typographical simplicity. Thus the scalar product of two vectors is simply indicated by their juxtaposition, no unnecessary dots, parentheses, commas, or brackets. Also, we do not make the distinction between row and column vectors, though this seems still to be the vogue in many quarters, for what reason we cannot imagine. Perhaps we are carrying typographical economy too far when we denote the vector x with coordinates from ξ_1 to ξ_n by the symbol (ξ_i) instead of the conventional $(\xi_1, \ldots, \xi_n)$, but why not? After all, nobody objects to indicating a matrix A in terms of its coordinates by the symbol (α_{ij}). We have gone to considerable length to avoid hanging subscripts on subscripts. The general philosophy has been that a clean-looking page of symbols will have a good psychological effect on the reader, or to put it the other way, a tangled symbolism suggests a tangled argument and is likely to frighten rather than entice.

About the most radical innovation in terminology is the replacement of the universally used "nonsingular" by "regular" in describing a square matrix of maximal rank. We just didn't like the sound of the double negative. Vector spaces have a certain "rank" rather than "dimension" simply because there is no reason to use two words for the same thing. "Polyhedral cone" hasn't been around very long yet. Perhaps we can persuade others to join us in calling them "finite cones." It does sound better, and as Professor Coxeter has pointed out, "polyhedron" belongs to the 3-space just as "polygon" belongs to the plane. The correct n-dimensional word is "polytope," and this is the word that will be encountered here.

Our system for numbering displayed relations is admittedly unorthodox. In each proof we start numbering the relations from the beginning starting with (1). Thus, if we argue that a certain conclusion follows from (3) we are referring to (3) in that same proof.

If the reader disagrees with some of the liberties we have taken we hope he will simply attribute them to temperament and forgive us. To ensure against the possibility of serious confusion we have included a table of notations at the front of the book and an index of terms at the back.

Finally, a word concerning the bibliography. We have listed conscientiously at the end of each chapter all sources which were actually used in its preparation. We have, however, made no attempt at bibliographical completeness, as this is not generally done in textbooks. The people whose names appear in the bibliography at the end of the book represent but a fraction of those who have made significant contributions to the subject—an ever-dwindling fraction since new investigators are constantly entering the field. For the reader who is interested in bibliographical matters we recommend the very complete "Bibliography on Linear Programming and Related Techniques" by Riley and Gass (Johns Hopkins Press, Baltimore, 1958).

Acknowledgments

This book evolved in three distinct stages. The initial stage involved a course given in the academic year 1956–1957 to a group of graduate students in pure and applied mathematics at Brown University. From this course my assistant Edmund Eisenberg and I assembled a set of mimeographed notes which were made available to the public in a limited supply. This enterprise was carried out in part under a contract with the Logistics Branch of the Office of Naval Research, to which I am indebted not only for financial support but also for encouragement and interest in the project.

Because of the favorable response to the course notes I decided to expand them into a textbook. Most of this work was done while I was working as a consultant to the Mathematical Analysis Division of The RAND Corporation in 1957–1958. RAND not only supplied me with all the physical equipment needed for this operation but, even more

important, enabled me to do the work in the place having the highest concentration of contributors to the subject about which I was writing. For this stimulating atmosphere I am ultimately indebted to the United States Air Force, whose Project RAND contract has enabled The RAND Corporation to undertake its broad program of scientific research. Part of this volume was given limited circulation in three Project RAND research memoranda. I will not even attempt to list all the people at RAND who have helped me in one way or another on various portions of the exposition, but should like to give special thanks to J. D. Williams, head of the Mathematics Division, who made it possible for me to come to RAND.

The final stages of writing were completed at Brown University in the fall of 1958, again with the support of the Logistics Branch of the Office of Naval Research.

Finally, I should like to thank Professor E. Barankin of the University of California whose suggestions based on a critical reading of the mimeographed course notes led me to make fairly extensive revisions in my original organization of material.

To all the above groups and individuals let me convey my gratitude and express the hope that the finished product presented herewith will to some extent justify their support.

David Gale

List of Notations

Below are listed the principal mathematical notations used in this book. The notations are listed in the order in which they occur in the text.

$\alpha, \beta, \gamma, \ldots, \xi, \eta, \zeta$, and other Greek letters represent numerical quantities also referred to as *scalars*

$a, b, c, \ldots, x, y, z$, and other italic letters represent vector quantities

$x = (\xi_i)$ the vector whose ith coordinate is ξ_i

$b = (\beta_j)$ the vector whose jth coordinate is β_j

$y = (\eta_1, \ldots, \eta_n)$ the vector whose coordinates are $\eta_1, \ldots, \eta_n$

F^n the set of all n-vectors over the field F

R^n *n space*, the set of all real n-vectors

u, v the *unit vectors* all of whose coordinates are one

$u_i, (v_j)$ the *ith (jth) unit vector* whose ith (jth) coordinate is one and whose other coordinates are zero

λx product of scalar λ with vector x

ε symbol for set-theoretic membership, "is an element of"

xy *scalar product* of vectors x and y

$A = (\alpha_{ij})$ the *matrix* whose ijth coordinate is α_{ij}

$a_i = (\alpha_{i1}, \ldots, \alpha_{in})$ the ith *row vector* of the matrix A

$a^j = (\alpha_{1j}, \ldots, \alpha_{mj})$ the jth *column vector* of the matrix A

$xA, (Ay)$ the product of the matrix A with the vector $x(y)$

L linear subspace of a vector space

L^* orthogonal or *dual* subspace of L

$x \geqq 0$ vector x is *nonnegative*

$x \geq 0$ vector x is *semipositive*

$x > 0$ vector x is *positive*

$M = \{1, \ldots, m\}$, $N = \{1, \ldots, n\}$ the set of positive integers from 1 to m and 1 to n, respectively

$\subset, \supset$ set-theoretic inclusion, "is contained in" and "contains," respectively

$\{x|P\}$ the set of all x such that x has property P

$\cup$ set-theoretic *union*

$\cap$ set-theoretic *intersection*

C convex cone

$C_1 + C_2$ algebraic *sum* of convex cones

C^* *dual cone* of C

P the *positive orthant*, all nonnegative vectors

(b) the *halfline* generated by the vector b

$(b)^*$ the *halfspace* generated by the vector b

$(a_1) + \cdots + (a_m)$ the finite cone generated by $a_1, \ldots, a_m$

K convex set

$<X>$ the *convex hull* of the set X

$<x_1, \ldots, x_n>$ the convex hull of vectors $x_1, \ldots, x_n$

$I = (\delta_{ij})$ the *identity matrix*

A^{-1} the *inverse* of the matrix A

A^* the *transpose* of the matrix A

$x \succ 0$ x is *lexicographically positive*

(N, k) *capacitated network* with *nodes* N and *capacity function* k

(x, y) *edge* from node x to node y

$g(A)$ values of function on nodes A of N given by $g(A) = \sum_{x \varepsilon A} g(x)$

$h(A, B)$ value of function on edges from A to B given by $h(A, B) = \sum_{x \varepsilon A, y \varepsilon B} h(x, y)$

s *source* in a network

s' *sink* in a network

(S, S') a *cut* in a network

∞ symbol for infinity

Γ two-person zero-sum game

P_1, P_2 first and second player

s, t *strategies* for first and second players

S, T *strategy sets* for first and second players

$(S, T; \phi)$ game with strategy sets S and T and *payoff* ϕ

σ, τ *mixed strategies* for first and second players

$<S>, <T>$ sets of mixed strategies for first and second players

$\bar{\sigma}, \bar{\tau}$ *optimal* mixed strategies

ω *value* of a game

$(\bar{\sigma}, \bar{\tau}; \omega)$ *solution* of a game in mixed strategies

$\bar{x}, \bar{y}$ *optimal strategies* for a matrix game

$U_m = \{u_1, \ldots, u_m\}$ set of *pure strategies* for a matrix game

$<U_m> = <u_1, \ldots, u_m>$ set of *mixed strategies* for a matrix game

$\bar{X}, \bar{Y}$ set of *optimal strategies* for a matrix game

$(\bar{X}, \bar{Y}; \omega)$ *solution* of a matrix game

$x_k \to x$ sequence x_k *converges* to x

$|x|$ *norm* of x

Contents

CHAPTER ONE

Linear Programming: Examples, Definitions, and Statement of the Principal Theorems

Maximum and minimum problems occur frequently in many branches of pure and applied mathematics. In economic applications such problems are especially natural. Firms try to maximize profits or minimize costs. Social planners attempt to maximize the welfare of the community. Consumers wish to spend their income in such a way as to maximize their satisfaction.

Linear programming is concerned with special classes of maximum and minimum problems which come up very frequently in economic applications. It is our purpose in this chapter to describe and define these problems in a precise manner. We shall then present the main theoretical results concerning them. The proofs of the results will be given in Chap. 3 after we have developed the necessary algebraic machinery in the next chapter.

It will be our policy here and throughout the book to introduce general concepts by means of concrete examples. Accordingly, the next section will be devoted to discussion of some specific instances of linear programs which will serve to guide us in formulating the general definitions which follow.

1. Examples

Example 1. The Diet Problem. This problem has become the classical illustration in linear programming and is treated in virtually every exposition of the subject. It is concerned with the problem of feeding, say, an army in the most economical way while at the same time satisfying certain nutritional requirements. Let us be specific. A dietitian is confronted with n different foods which will be labeled $F_1, F_2, \ldots, F_n$. From these he is to select a *diet*, that is, he must determine the amount of each food which is to be consumed annually by a person or group of persons. This yearly menu is required to supply certain amounts of various nutritional elements such as proteins, calories, minerals, vitamins, and the like. We shall refer to these types of nutritive elements simply as *nutrients* of which there will be m varieties denoted by $N_1, \ldots, N_m$. We suppose that each man is required to consume at least γ_1 units of N_1, γ_2 units of $N_2, \ldots,$ γ_m units of N_m, per year. In order to meet these requirements the dietitian must know exactly how much of each nutrient is contained in each of the foods. Let us denote by α_{ij} the amount of the ith nutrient contained in one unit of the jth food. The information which the dietitian needs is then conveniently presented in the following table or *matrix*:

	F_1	F_2	$\cdots$	F_n
N_1	α_{11}	α_{12}		α_{1n}
N_2	α_{21}	α_{22}		α_{2n}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
N_m	α_{m1}	α_{m2}		α_{mn}

The entry in the ith row and the jth column of the matrix is the number α_{ij}, giving the amount of N_i in one unit of F_j. We shall refer to the table above as the *nutrition matrix* of the problem.

Suppose now that the dietitian has chosen a diet. This means that he has determined that η_1 units of F_1, η_2 units of F_2, etc., shall be consumed per man per year. How does he now check that the nutritional requirements are satisfied by this diet? Obviously, he simply calculates the amount of each nutrient in the diet and compares it with the prescribed amount. Consider the nutrient N_1. Each unit of F_1 contains α_{11} units of N_1, and since there are η_1 units of F_1 in the diet, we get $\eta_1\alpha_{11}$ units of N_1 from F_1. Similarly we get $\eta_2\alpha_{12}$ units of N_1

from F_2, and in general $\eta_j\alpha_{1j}$ units of N_1 from F_j. The total amount of N_1 in this diet is then

$$\eta_1\alpha_{11} + \eta_2\alpha_{12} + \cdots + \eta_n\alpha_{1n}$$

and this amount is required to be at least equal to γ_1. Thus the requirement on N_1 simply states that the numbers $\eta_1, \ldots, \eta_n$ must satisfy the inequality

$$\sum_{j=1}^{n} \eta_j\alpha_{1j} \geqq \gamma_1$$

The requirements on the remaining nutrients take exactly the same form, and the condition that the diet satisfy all requirements is that the numbers η_j satisfy simultaneously the m inequalities

$$\sum_{j=1}^{n} \eta_j\alpha_{ij} \geqq \gamma_i \qquad \text{for } i = 1, 2, \ldots, m \tag{1}$$

A diet for which conditions (1) are satisfied will be termed a *feasible diet.*

As yet no maximum or minimum problem has been described, but we have already mentioned that the dietitian must choose the most economical diet consistent with the requirement (1). We are assuming then that a *price* is associated with each food. Let π_j be the price of one unit of the food F_j. It follows that the cost of the diet described by the numbers η_j is given by the expression

$$\pi_1\eta_1 + \pi_2\eta_2 + \cdots + \pi_n\eta_n = \sum_{j=1}^{n} \pi_j\eta_j \tag{2}$$

We can now give a complete statement of the diet problem. *Among all diets satisfying conditions* (1) *find one such that expression* (2) *is a minimum.*

The problem which we have just described in perhaps tedious detail is a typical linear programming problem. The word "linear" is used because both the inequalities (1) and the function to be minimized (2) are linear.

A diet which satisfies both (1) and (2) is called an *optimal* diet. Mathematically the diet problem can be broken into two parts: first, that of finding a feasible diet and, second, if a feasible diet exists, of finding an optimal diet. It is easy to see that a feasible diet will always exist provided each nutrient N_i occurs in at least one food F_j, for then by using a sufficient amount of the foods one can always satisfy the requirements. It is quite clear in this case that an optimal diet also exists. A rigorous proof of this fact will have to wait, however, until a later chapter.

Example 2. The Transportation Problem. Let a certain commodity, say steel, be produced at each of m plants, $P_1, \ldots, P_m$, and let σ_i (σ = supply) be the yearly output of the ith plant. Suppose now that steel is required at each of n markets, $M_1, \ldots, M_n$, and let the annual demand at the jth market be δ_j. Finally, let γ_{ij} be the cost of shipping one unit from P_i to M_j.

The problem is now to determine a *shipping schedule* such that (1) the demand δ_j at the market M_j will be satisfied, (2) the supply σ_i at the plant P_i will not be exceeded, and (3) the total shipping cost will be a minimum. A shipping schedule consists simply of mn nonnegative numbers ξ_{ij}, where ξ_{ij} represents the amount to be shipped from P_i to M_j. The total amount shipped into M_j is thus $\sum_{i=1}^{m} \xi_{ij}$, and condition (1) becomes

$$\sum_{i=1}^{m} \xi_{ij} \geqq \delta_j \tag{1}$$

The total amount shipped out of P_i is $\sum_{j=1}^{n} \xi_{ij}$, and condition (2) is therefore

$$\sum_{j=1}^{n} \xi_{ij} \leqq \sigma_i \tag{2}$$

and, finally, we are required to minimize

$$\sum_{i,j} \gamma_{ij}\xi_{ij} \tag{3}$$

It will be noted that this problem is of the same general form as the diet problem. We are seeking certain nonnegative numbers ξ_{ij} which satisfy the system of *linear inequalities* (1) and (2) and minimize the *linear function* (3). In analogy with terminology used in the diet problem, we shall say that a shipping schedule is *feasible* if the numbers ξ_{ij} satisfy inequalities (1) and (2). It is immediately clear that a necessary condition for the problem to be feasible is the requirement that total supply be at least as large as the total demand, that is,

$$\sum_{i=1}^{m} \sigma_i \geqq \sum_{j=1}^{n} \delta_j \tag{4}$$

Conversely, we leave it to the reader to prove that if (4) is satisfied then there exists a feasible shipping schedule (see Exercise 1).

The next two examples are of quite general importance and include many others as special cases. These involve the important idea of a *linear production model*, which we now describe. Consider a production system, say a factory, in which n goods $G_1, \ldots, G_n$ are involved either as inputs to the productive process or as final goods. For instance, these goods might include steel, labor, automobiles, etc. Goods are produced by *linear processes* or *activities* which are to be thought of as "recipes" giving the proportions of the various goods required in a given mode of production. An ordinary cookbook recipe provides a typical example. Thus, if the goods are, say, eggs, butter, salt, milk, cheese, the *soufflé activity* is completely described by stating how many parts of butter, eggs, etc., are required to produce one unit of soufflé. The statement that this process is *linear* simply means that multiplying all ingredients by any constant multiplies the amount of soufflé by the same constant. Note that this assumption of linearity is a rather severe restriction on the types of processes to be considered. If in the culinary example above we had included labor among the goods, the linearity feature would have been lost, for it is certainly not true in general that preparing a double portion of soufflé requires twice the cooking labor.

A formal definition of an activity is now easily given.

Definition. An *activity* P involving n goods corresponds to a set of n numbers, $\alpha_1, \ldots, \alpha_n$. The good G_j is called an *input* to the activity if α_j is negative, and an *output* of the activity if α_j is positive.

A *linear production model* P involving n goods consists of a set of such activities $P_1, \ldots, P_m$. Such a model is completely described by an array of mn numbers α_{ij}, where α_{ij} is the amount of G_j produced (or consumed if α_{ij} is negative) when P_i is operated at unit level. This array of numbers is called the *production matrix* of the model.

$$\begin{array}{c|ccc} & G_1 & & G_n \\ \hline P_1 & \alpha_{11} & \cdots & \alpha_{1n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ P_m & \alpha_{m1} & \cdots & \alpha_{mn} \end{array}$$

Now, in order to describe completely how the production model behaves it is necessary to specify the inputs and outputs of each of the activities. We shall say the activity P_i is being *operated at the level* or *intensity* ξ_i if its inputs and outputs are given by the numbers $\xi_i\alpha_{i1}, \xi_i\alpha_{i2}, \ldots, \xi_i\alpha_{in}$. A *production schedule* for P is defined to be a set of nonnegative intensities $\xi_1, \ldots, \xi_m$ for the activities P_i. Given these numbers ξ_i we see that the total amount of G_j produced is the sum of the amounts produced by each of the activities and is given by the expression

$$\xi_1\alpha_{1j} + \xi_2\alpha_{2j} + \cdots + \xi_m\alpha_{mj}$$

where, of course, this quantity may be negative, which simply means that the jth good is being consumed rather than produced.

We are now prepared to present the two examples.

Example 3. Production to Meet Given Demand at Minimum Cost. Assume we have a linear production model and it is required that we produce at least δ_j units of G_j (δ = demand). Suppose further that the cost of operating the process P_i at unit level is γ_i; hence the cost of operating P_i at level ξ_i is $\xi_i\gamma_i$, where we are again making the rather restrictive assumption that the cost of operating an activity is proportional to the level at which it is operated. The problem is then to choose a production schedule which will satisfy the demands δ_j and minimize the total cost. Thus, we seek nonnegative numbers $\xi_1, \ldots, \xi_m$ which

$$\text{minimize} \sum_{i=1}^{m} \xi_i\gamma_i \qquad (1)$$

subject to the requirements that

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \geqq \delta_j \qquad \text{for } j = 1, \ldots, n \tag{2}$$

Some remarks are in order here. Recall that the numbers α_{ij} may be either positive or negative according as G_j is produced or consumed by the process P_i. Similarly, the demands δ_j may be negative. A negative demand is economically a supply, for if δ_j is negative, the inequality

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \geqq \delta_j$$

means that we must not consume more than the amount $-\delta_j$. Thus, both supplies and demands are taken account of in this model.

The question of feasibility is no longer a simple one in this example. It may easily happen that it is not technologically possible to satisfy the given demands with the given resources, for (2) may be any system of linear inequalities and such systems need not have solutions. If, however, a feasible schedule does exist, it is intuitively clear that there is an *optimal* schedule since the cost of a given schedule is bounded below by zero. A proof of this fact will be given in Chap. 3.

Example 4. Production to Maximize Income from Given Resources. This example is essentially the same as the previous one except for a change of sign. Again we consider a linear production model but instead of associating a cost with each activity we let $\gamma_i \geqq 0$ be the rate of *return* or *income* associated with the activity P_i, obtained, say, from selling the outputs of the activity. Assume further that there is a given fixed supply σ_j of the jth good. The problem is now to find a production schedule $\xi_1, \ldots, \xi_m$ which will maximize the total income without exceeding the given supplies. In symbols, we wish to

$$\text{maximize} \sum_{i=1}^{m} \xi_i \gamma_i \tag{1}$$

subject to the conditions

$$-\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \sigma_j \tag{2}$$

The reason for the negative sign in (2) is that we are here taking supplies as positive and thus the amount of G_j consumed in production is not to exceed σ_j.

With the four examples before us it is not difficult to see what elements they have in common, and we can give our first basic definition.

Definition. A *standard linear programming problem* is that of finding nonnegative numbers $\xi_1, \ldots, \xi_m$ which either maximize or minimize a given linear function, that is,

$$\sum_{i=1}^{m} \xi_i \gamma_i \textit{ is to be maximum or minimum} \tag{1}$$

where the numbers ξ_i *are also required to satisfy a set of linear inequalities.*

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \beta_j \qquad j = 1, \ldots, n \tag{2}$$

We shall later define a more general kind of linear program. For the present the reader should observe that all our examples fall under the definition above. In accordance with previous terminology we shall call numbers ξ_i which satisfy (2) a *feasible solution* of the problem, and we shall call a problem *feasible* if it has a feasible solution. A feasible solution which maximizes or minimizes (1) will be called an *optimal solution.* The number giving the maximum or minimum will then be called the *value* of the linear program.

2. Duality and Prices

We begin this section by considering a specific maximum problem.

Example 5. Find nonnegative numbers ξ_1, ξ_2, ξ_3, ξ_4 such that

$$2\xi_1 + 4\xi_2 + \xi_3 + \xi_4 \text{ is a maximum} \tag{1}$$

subject to the conditions

$$\begin{aligned} \xi_1 + 3\xi_2 \qquad\quad + \xi_4 &\leqq 4 \\ 2\xi_1 + \xi_2 \qquad\qquad\qquad &\leqq 3 \\ \xi_2 + 4\xi_3 + \xi_4 &\leqq 3 \end{aligned} \tag{2}$$

We assert: An optimal solution of this problem is given by

$$\xi_1 = 1 \qquad \xi_2 = 1 \qquad \xi_3 = \tfrac{1}{2} \qquad \xi_4 = 0$$

The reader will verify by direct substitution that these numbers are feasible; i.e., they satisfy the inequalities (2) above. Substituting the numbers in (1) we obtain

$$2 \cdot 1 + 4 \cdot 1 + \tfrac{1}{2} + 0 = 6\tfrac{1}{2}$$

and it is our claim that $6\frac{1}{2}$ is in fact the desired maximum. How do we know this and how can we be sure that some other choice of the numbers ξ_i will not give us a larger value of (1) and still satisfy (2)? In the paragraphs that follow we are going to prove to the reader that the feasible solution above is actually optimal. In order to do this we turn for a moment to the general problem of finding *nonnegative numbers* $\xi_1, \ldots, \xi_m$ *which*

$$\textit{maximize} \sum_{i=1}^{m} \xi_i \gamma_i \tag{3}$$

subject to the inequalities

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \beta_j \qquad j = 1; \ldots, n \tag{4}$$

The fundamental fact about linear programming is that to the maximum problem above corresponds the following standard minimum problem: *find nonnegative numbers* $\eta_1, \ldots, \eta_n$ *which*

$$\textit{minimize} \sum_{j=1}^{n} \eta_j \beta_j \tag{3)*}$$

subject to the inequalities

$$\sum_{j=1}^{n} \eta_j \alpha_{ij} \geqq \gamma_i \qquad i = 1, \ldots, m \tag{4)*}$$

Problem (3)*, (4)* is called the *dual* of problem (3), (4) and the central results of linear programming theory concern the relationship between a problem and its dual. We shall shortly give a precise statement of this relationship. At this time we make the following observation.

Lemma 1.1. *Let* $\xi_1, \ldots, \xi_m$ *be a feasible solution of a standara maximum problem* [thus, a nonnegative solution of inequalities (4)] *and let* $\eta_1, \ldots, \eta_n$ *be a feasible solution of the dual problem* [a nonnegative solution of inequalities (4)*]. *Then*

$$\sum_{i=1}^{m} \xi_i\gamma_i \leqq \sum_{i,j} \xi_i\eta_j\alpha_{ij} \leqq \sum_{j=1}^{n} \eta_j\beta_j \tag{5}$$

Proof. Multiplying the jth inequality of (4) by η_j and summing on j gives

$$\sum_{j=1}^{n} \eta_j\beta_j \geqq \sum_{j=1}^{n} \eta_j \sum_{i=1}^{m} \xi_i\alpha_{ij} = \sum_{i,j} \xi_i\eta_j\alpha_{ij} \tag{6}$$

Multiplying the ith inequality of (4)* by ξ_i and summing on i gives

$$\sum_{i=1}^{m} \xi_i\gamma_i \leqq \sum_{i=1}^{m} \xi_i \sum_{j=1}^{n} \eta_j\alpha_{ij} = \sum_{i,j} \xi_i\eta_j\alpha_{ij} \tag{7}$$

and (6) and (7) together yield (5).

As a consequence of this lemma we have our first important result.

Theorem 1.1 (optimality criterion). *If there exist feasible solutions* $\xi_1, \ldots, \xi_m$ *and* $\eta_1, \ldots, \eta_n$ *for the maximum problem above and its dual such that*

$$\sum_{i=1}^{m} \xi_i\gamma_i = \sum_{j=1}^{n} \eta_j\beta_j \tag{8}$$

then these feasible solutions are, in fact, optimal solutions of their respective problems.

Proof. Let $\xi'_1, \ldots, \xi'_m$ be any other feasible solution of the maximum problem. Then from the lemma

$$\sum_{i=1}^{m} \xi'_i\gamma_i \leqq \sum_{j=1}^{n} \eta_j\beta_j \tag{9}$$

and combining this with (8) gives

$$\sum_{i=1}^{m} \xi'_i\gamma_i \leqq \sum_{i=1}^{m} \xi_i\gamma_i$$

showing that $\xi_1, \ldots, \xi_m$ is an optimal solution. A symmetrical argument proves the optimality of $\eta_1, \ldots, \eta_n$.

Let us immediately apply this result to our numerical problem. The dual of this problem is seen to be that of finding nonnegative numbers η_1, η_2, η_3 such that

$$4\eta_1 + 3\eta_2 + 3\eta_3 \text{ is a minimum subject to} \qquad (1)^*$$

$$\begin{aligned} \eta_1 + 2\eta_2 \qquad\quad &\geqq 2 \\ 3\eta_1 + \eta_2 + \eta_3 &\geqq 4 \\ 4\eta_3 &\geqq 1 \\ \eta_1 + \qquad \eta_3 &\geqq 1 \end{aligned} \qquad (2)^*$$

Now, one verifies by direct substitution that

$$\eta_1 = {}^{11}\!/_{10} \qquad \eta_2 = {}^{9}\!/_{20} \qquad \eta_3 = {}^{1}\!/_{4}$$

provides a nonnegative solution of (2)*, hence a feasible solution of the dual problem. Furthermore, evaluating (1)* gives

$$4 \cdot {}^{11}\!/_{10} + 3 \cdot {}^{9}\!/_{20} + 3 \cdot {}^{1}\!/_{4} = {}^{130}\!/_{20} = 6{}^{1}\!/_{2}$$

which is the same as the value obtained from the feasible solution of the original maximum problem. It follows from Theorem 1.1 that we have found optimal solutions for both the original problem and its dual.

We have now fulfilled our promise of proving to the reader that the feasible solution which we exhibited at the beginning of this section is optimal. This was possible because we were able to find a feasible solution of the dual problem which together with the original solution satisfied the optimality criterion. A natural question which has now perhaps occurred to the reader is this: Was it just a fortunate accident that we were able to find a suitable solution of the dual problem in this case, or can we expect such solutions to exist in general? The central fact in the theory of linear programming is that the phenomenon noted in this example holds for all linear programming problems. In precise terms the converse of Theorem 1.1 is also true; if we have optimal solutions of a problem and its dual then the values of the two problems are equal. This result is known as the fundamental duality theorem of linear programming, which we state as follows:

Fundamental Duality Theorem. *If a standard maximum or minimum problem and its dual are both feasible then they both have optimal solutions and both have the same value. If either problem is not feasible then neither has an optimal solution.*

This rather remarkable result, which seems to have been noted first by von Neumann, is basic not only in the theory of linear programming but also in two-person game theory and a number of other branches of linear economic theory. The proof is not simple and will have to be put off until the necessary algebraic machinery has been developed in the next chapter. We shall devote the rest of this chapter to interpreting the result in economic terms and to deducing some of its consequences.

In order to gain further understanding of the significance of the duality theory let us return to the numerical problem of Example 5 and interpret this example as a production problem of the type described in Example 4. We assume we have 4 activities P_1, P_2, P_3, and P_4. The income from operating P_1 at unit level is 2, that for P_2 is 4, etc. There are 3 goods G_1, G_2, and G_3 and the activity P_1 requires as inputs 1 unit of G_1 and 2 units of G_2, while P_2 requires 3 units of G_1, 1 unit of G_2, and 1 unit of G_3, etc. There are available 4 units of G_1, 3 units of G_2, and 3 units of G_3. At what levels shall the processes be operated so as to maximize the total income?

We have given a typical interpretation of the problem (1), (2), but how are we to interpret the dual problem (1)*, (2)*? First note that the right-hand side of the inequalities (2)* is income and is therefore measured in monetary units, say in dollars. The coefficients on the left-hand side of (2)* are in units of goods. It follows that the numbers η_j have the units of dollars per unit of goods, that is, the η_j are *unit prices* of the goods G_1, G_2, and G_3. With this interpretation of the unknowns η_j what is the meaning of the inequalities (2)*? Let us consider the first inequality

$$1 \cdot \eta_1 + 2 \cdot \eta_2 \geqq 2$$

The coefficients 1 and 2 on the left are the amounts of G_1 and G_2 required to operate activity P_1 at unit level. Since η_1 and η_2 are the corresponding prices the left-hand side above is the *cost* of operating

P_1 at unit level, and the inequality states that this cost must be at least as great as the income received from the activity. The same analysis applies to the other inequalities (2)*, whose meaning can now be summarized by the simple economic statement:

The prices η_j must be such that no activity P_i makes a positive profit.

Finally (1)* is seen to be the requirement that the total value of the resources be minimized. It is now possible to give a verbal argument justifying Theorem 1.1. It runs like this: We have found prices η_j with the property that the return from each activity will be no greater than the cost of the activity. Therefore, the total return from operating the model is at most equal to the total cost of the available resources. But we have found a way of operating the model in which the return is equal to this total cost, and therefore this mode of operation must be optimal.

3. Further Interpretation of Duality

We have now stated the fundamental duality theorem and shown how it can be used to prove that a given feasible solution of a program is optimal. In this section we shall look again at the examples of linear programs given in Sec. 1 and interpret the duality theorem in each case.

1. The Diet Problem. We shall not rewrite the problem formally but simply remind the reader that it concerned selecting a diet satisfying certain requirements on nutrients and minimizing total cost. The dual problem can be described as follows: to assign values or "prices" to each nutrient N_i in such a way that the sum of the values of the nutrients in one unit of the jth food F_j does not exceed its unit cost π_j, and such that the total value of the amounts of nutrients required by the diet is a maximum. It is recommended that the reader verify formally the above verbal statement.

An important nonmathematical question now arises. Is there any way of looking at this dual problem so that it makes sense economically? It is quite clear why a dietitian would want to minimize the cost of an adequate diet, but why would anyone want to maximize the value of the nutrients in such a diet? We are about to describe a simple situation in which such a maximization has an economic meaning. The description may seem artificial at first, but it will appear less so as we go on to consider other examples. The situation requires that we introduce a new character into our gastronomic drama, a seller of pills—vitamin pills, iron capsules, and so forth. This salesman is, in fact, able to provide the dietitian with all the nutrients the latter needs in some concentrated form. The dietitian, whose sole aim is to minimize costs, will willingly substitute pills and capsules for steak and potatoes provided this will save money (a certain lack of realism in the original problem is becoming increasingly apparent, but this is, of course, beside the point for present purposes). Suppose then that the pill salesman sets the prices of a unit of N_i (the ith nutrient) at some value ξ_i, making sure that

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \pi_j \qquad \text{for all } j, \text{ where } \alpha_{ij} \text{ is the amount of } N_i \text{ in } F_j \qquad (2)^*$$

This means that the total value of the nutrients in a unit of F_j is no greater than the unit cost of F_j. It is now clear that, no matter what diet he chooses, it will always be at least as economical for the dietitian to buy pills since the cost of each food is at least as great as the cost of the nutrients it contains. The pill man will, however, now charge the dietitian as much as possible subject to the constraints (2)*. Since the adequate diet calls for γ_i units of N_i, he sets prices ξ_i so as to maximize

$$\sum_{i=1}^{m} \xi_i \gamma_i \qquad (1)^*$$

and this is precisely the dual problem.

We can be somewhat less concrete in our description of the dual problem by saying that the nutrient prices ξ_i are those which enable the pill man to realize the maximum return and still compete favorably with the grocer. It is this idea of *competitive* prices which is characteristic of the interpretation of the duality theorem.

2. The Transportation Problem. In order to determine the dual here it is convenient to write out the relations without the summation symbols. Recall that we wish to choose numbers ξ_{ij} which minimize the shipping cost

$$\begin{array}{l} \xi_{11}\gamma_{11} + \xi_{12}\gamma_{12} + \cdots + \xi_{1n}\gamma_{1n} \\ + \xi_{21}\gamma_{21} + \xi_{22}\gamma_{22} + \cdots + \xi_{2n}\gamma_{2n} \\ + \cdots\cdots\cdots\cdots\cdots \\ + \xi_{m1}\gamma_{m1} + \xi_{m2}\gamma_{m2} + \cdots + \xi_{mn}\gamma_{mn} \end{array} \tag{1}$$

subject to the inequalities (see Sec. 1, Example 2)

$$\begin{array}{lllr} -\xi_{11} - \cdots - \xi_{1n} & & & \geqq -\sigma_1 \\ & - \xi_{21} - \cdots - \xi_{2n} & & \geqq -\sigma_2 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots \\ & & - \xi_{m1} - \cdots - \xi_{mn} & \geqq -\sigma_m \\ \xi_{11} & + \xi_{21} + \cdots & + \xi_{m1} & \geqq \delta_1 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots \\ \quad \xi_{1n} & \quad + \xi_{2n} + \cdots & \quad + \xi_{mn} & \geqq \delta_n \end{array}$$

where σ_i is the supply at plant P_i and δ_j is the demand at market M_j.

The dual problem now becomes: to determine nonnegative numbers π_i, $i = 1, \ldots, m$, and π'_j, $j = 1, \ldots, n$, such that $\pi'_j - \pi_i \leqq \gamma_{ij}$ for all i and j, and such that $\sum_{j=1}^{n} \pi'_j\delta_j - \sum_{i=1}^{m} \pi_i\sigma_i$ is a maximum. The reader should verify that this is the correct statement of the dual problem.

How shall we interpret this dual problem? First, notice that since we have the relations $\pi'_j - \pi_i \leqq \gamma_{ij}$ we are forced once again to measure the variables π'_j and π_j in units of money. Let us consider now that the numbers γ_{ij} represent the established transportation costs which confront, say, a steel manufacturer who is in the act of trying to decide on a shipping schedule. He is interrupted by a visit from a representative of the new Fly-By-Nite Transportation Company, who makes him the following proposition: "I will buy all your steel, paying π_i for each unit of steel at plant P_i. I will guarantee to deliver the steel to the markets M_j in the quantities δ_j and I will then sell it back to you, charging π'_j for each unit at M_j. Please notice that

$$\pi'_j - \pi_i \leqq \gamma_{ij} \qquad \text{for all } i \text{ and } j \tag{2}^*$$

so you will pay no more than you would if you paid the normal transportation costs." The manufacturer is forced to agree on this point and the deal is therefore closed to the satisfaction of the Fly-By-Nite man as well, for he has shrewdly set the prices so as to maximize his profit $\Sigma\pi'_j\delta_j - \Sigma\pi_i\sigma_i$ subject only to $(2)^*$ above.

Because of the duality theorem, it will turn out that the manufacturer doesn't actually save any money by this maneuver (though he is saved the trouble of calculating the minimal shipping schedule).

Let us illustrate the duality theory for the transportation problem by a numerical example.

Example 6. The diagram in Fig. 1.1 gives a schematic representation of a transportation problem with two plants and three markets.

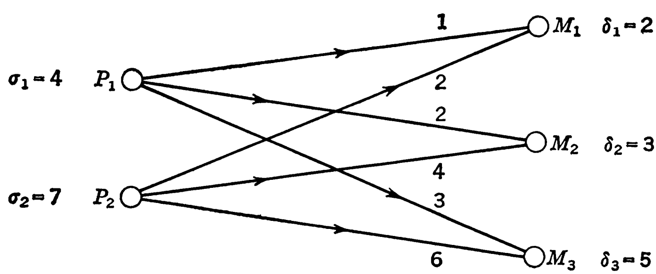


FIG. 1.1

This graphical representation is almost self-explanatory. The vertices of the graph represent the plants and markets, and the corresponding supplies σ_i and demands δ_j are indicated. The lines connecting plants and markets represent the various routes and the number above each line gives the cost of the corresponding route. The *cost matrix* is seen to be

	M_1	M_2	M_3
P_1	1	2	3
P_2	2	4	6

where the ijth entry above is the unit shipping cost from P_i to M_j.

We now claim that a solution to the above problem is given by

$$\xi_{11} = 0 \qquad \xi_{12} = 0 \qquad \xi_{13} = 4$$
$$\xi_{21} = 2 \qquad \xi_{22} = 3 \qquad \xi_{23} = 1$$

and the minimum cost is given by

$$0 \cdot 1 + 0 \cdot 2 + 4 \cdot 3 + 2 \cdot 2 + 3 \cdot 4 + 1 \cdot 6 = 12 + 4 + 12 + 6 = 34$$

In order to prove that this is a minimum, we exhibit the following feasible prices for the dual problem:

$$\begin{array}{lll} \pi_1 = 3 & \pi_2 = 0 & \\ \pi'_1 = 2 & \pi'_2 = 4 & \pi'_3 = 6 \end{array}$$

To verify feasibility, we must check that $\pi'_j - \pi_i$ is not greater than the ijth entry in the cost matrix. The following figure will facilitate this verification:

	$\pi'_1 = 2$	$\pi'_2 = 4$	$\pi'_3 = 6$
$\pi_1 = 3$	1	2	3
$\pi_2 = 0$	2	4	6

One easily checks that each entry in the cost matrix is not less than the difference between the numbers at the head of its column and row.

Finally we see that our solutions of the dual problems are optimal for we compute

$$\Sigma\pi'_j\delta_j - \Sigma\pi_i\sigma_i = 2 \cdot 2 + 4 \cdot 3 + 5 \cdot 6 - 3 \cdot 4 - 0 \cdot 7 = 34$$

and since this is the same as the shipping cost computed above we have solved the problem.

The solution can also be written down in tabular form as a *shipping matrix*

0	0	4
2	3	1

where the ijth entry is the amount shipped from P_i to M_j. Comparing the shipping matrix with the cost matrix we note that in our solution the cheapest route, that from P_1 to M_1, is not used, while the most expensive route, that from P_2 to M_3, is used. This fact is perhaps somewhat surprising and shows that it is not easy to guess the solution of a transportation problem in advance.

3. Production to Meet Given Demand at Minimum Cost. The interpretation of the dual here is very similar to that of the diet problem. The details are left to the reader.

4. Production to Maximize Return from Given Resources. This is the problem treated in the previous section. The dual consists in assigning prices π_j to the various goods G_j in such a way that

$$\sum_{j=1}^{n} \pi_j \sigma_j \text{ is minimized} \tag{1)*$$

subject to

$$-\sum_{j=1}^{n} \pi_j \alpha_{ij} \geqq \gamma_i \tag{2)*$$

Interpretation. A competitor believes he has a more efficient way of utilizing the given resources and wants to buy out the producer. He therefore offers to pay the producer the amount π_j for each unit of G_j, where the numbers π_j satisfy (2)* above. The competitor quickly convinces the producer that the amount of money offered is at least as much as he could obtain from any production schedule, "for," says the competitor, "if you operate P_i at level ξ_i your return will be

$$\sum_{i=1}^{m} \xi_i \gamma_i \tag{3}$$

where, of course, because of your limited supplies,

$$-\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \sigma_j \tag{4}$$

But if you sell to me, your return will be

$$\sum_{j=1}^{n} \pi_j \sigma_j$$

and

$$\begin{aligned} \sum_j \pi_j \sigma_j &\geqq -\sum_j \pi_j \sum_i \xi_i \alpha_{ij} && \text{[from (4)]} \\ &= -\sum_i \xi_i \sum_j \pi_j \alpha_{ij} \geqq \sum_i \xi_i \gamma_i && \text{[from (2)*]} \end{aligned} \tag{5}$$

so you will be at least as well off." Conclusion: The producer accepts the offer and the competitor buys him out at the lowest possible figure [i.e., condition (1)* subject to (2)*].

4. *Price Equilibrium*

In this section we shall present a result which is a rather simple consequence of the duality theorem. It is, however, of considerable economic importance, providing our first example of an "equilibrium theorem," a terminology which will be explained shortly.

We return to consideration of the standard maximum problem of finding nonnegative numbers $\xi_1, \ldots, \xi_m$ which

$$\text{maximize} \sum_{i=1}^{m} \xi_i \gamma_i \tag{1}$$

subject to

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \leqq \beta_j \qquad j = 1, \ldots, n \tag{2}$$

and the dual problem of finding nonnegative numbers $\eta_1, \ldots, \eta_n$ which

$$\text{minimize} \sum_{j=1}^{n} \eta_j \beta_j \tag{1*}$$

subject to

$$\sum_{j=1}^{n} \eta_j \alpha_{ij} \geqq \gamma_i \qquad i = 1, \ldots, m \tag{2*}$$

Assuming the duality theorem to be true we shall now prove the following result.

Theorem 1.2 (equilibrium theorem). *The feasible solutions* $\xi_1, \ldots, \xi_m$ *and* $\eta_1, \ldots, \eta_n$ *of* (2) *and* (2)*, *respectively, are optimal solutions if and only if*

$$\eta_j = 0 \qquad \textit{whenever} \qquad \sum_{i=1}^{m} \xi_i \alpha_{ij} < \beta_j \tag{3}$$

and

$$\xi_i = 0 \qquad \textit{whenever} \qquad \sum_{j=1}^{n} \eta_j \alpha_{ij} > \gamma_i \tag{3*}$$

Proof. First, suppose conditions (3) and (3)* hold. Multiplying the jth inequality of (2) by η_j and summing on j and making use of (3) gives

$$\sum_{j=1}^{n} \eta_j\beta_j = \sum_{j=1}^{n} \eta_j \sum_{i=1}^{m} \xi_i\alpha_{ij} = \sum_{i,j} \xi_i\eta_j\alpha_{ij} \tag{4}$$

Similarly, from (2)* and (3)* we get

$$\sum_{i=1}^{m} \xi_i\gamma_i = \sum_{i=1}^{m} \xi_i \sum_{j=1}^{n} \eta_j\alpha_{ij} = \sum_{i,j} \xi_i\eta_j\alpha_{ij} \tag{5}$$

and (4) and (5) show that

$$\sum_{i=1}^{m} \xi_i\gamma_i = \sum_{j=1}^{n} \eta_j\beta_j$$

whence from Theorem 1.1 the ξ_i and η_j are optimal solutions.

Conversely, if the ξ_i and η_j provide optimal solutions then from the duality theorem we know that

$$\sum_{i=1}^{m} \xi_i\gamma_i = \sum_{i,j} \xi_i\eta_j\alpha_{ij} = \sum_{j=1}^{n} \eta_j\beta_j$$

From the first equation we have

$$\sum_{i=1}^{m} \xi_i\left(\gamma_i - \sum_{j=1}^{n} \eta_j\alpha_{ij}\right) = 0$$

but since the numbers η_j are feasible it follows that the terms $\left(\gamma_i - \sum_{j=1}^{n} \eta_j\alpha_{ij}\right)$ are nonpositive and hence for each i

$$\xi_i\left(\gamma_i - \sum_{j=1}^{n} \eta_j\alpha_{ij}\right) = 0$$

from which (3)* follows at once. A symmetrical argument proves condition (3).

We shall now interpret the above result economically and justify the use of the word "equilibrium." For this purpose let us think of (1), (2) above as the production problem of Example 4. We have already seen that it is natural to interpret the dual variables η_j as prices, and we have also seen that the feasibility conditions (2)* correspond to the requirement that no activity makes a positive profit.

Condition (3)* has a very obvious interpretation. It says that if the cost of an activity exceeds the income derived from it then it will not be used; i.e., it will be operated at level zero. Conditions (2)* and (3)* together may be thought of as *stability conditions* in the following sense. If the model is operating at activity levels $\xi_1, \ldots, \xi_m$ and these conditions are satisfied then there will be no incentive to change the activity levels since there is no way of increasing income. Looked at the other way, if conditions (2)* or (3)* failed to hold then activity levels would be unstable, for the producer could increase his income by changing the production levels.

As to conditions (2) and (3), the first is simply the technological requirement that the available supply must not be exceeded. Condition (3) states that, if there are goods of which there is a *surplus*, that is, whose supply is not exhausted, then the price of these goods must be zero. This is also a stability condition, this time on prices rather than on activity levels. Recall that according to the classical "law of supply and demand" if the supply of a good exceeds the demand for it then its price will drop. On the other hand, prices cannot drop below zero and therefore a good which is oversupplied even when income is being maximized must become a *free good.*

As a second illustration of the equilibrium theorem, let us see what it says for the case of the transportation problem. Recall that a feasible shipping schedule is one which satisfies the given demands without exceeding the given supplies, and a feasible set of dual variables are prices at each plant and market with the property that the difference between market price and plant price does not exceed the shipping cost from plant to market. The equilibrium conditions now become:

(3)* If the difference between the price at a particular market and a particular plant is less than the corresponding shipping cost, then no goods will be shipped from that plant to that market.

Interpretation. "The company" will lose money if it costs more to ship from plant to market than what can be realized by sales at the market. Such unprofitable routes will not be used.

(3) If the amount shipped out of some plant is less than the supply at that plant then the price at that plant must be zero.

Interpretation. As in the previous discussion, if there is a surplus at some plant, the price there must drop to zero.

One of the most important uses of the equilibrium theorem is in connection with numerical computation. We have already seen that if feasible solutions of the primal and dual problems are given they can easily be checked for optimality. Now, using the equilibrium theorem we can often find the solution to the dual problem when the solution to the primal is given. Let us return to the numerical Example 5 in which the inequalities were

$$\begin{aligned} \xi_1 + 3\xi_2 \qquad\qquad + \xi_4 &\leqq 4 \\ 2\xi_1 + \xi_2 \qquad\qquad\qquad &\leqq 3 \\ \xi_2 + 4\xi_3 + \xi_4 &\leqq 3 \end{aligned} \tag{2}$$

and the proposed optimal solution was

$$\xi_1 = 1 \qquad \xi_2 = 1 \qquad \xi_3 = \tfrac{1}{2} \qquad \xi_4 = 0$$

According to the equilibrium theorem the dual inequalities must in fact be equations for the cases $i = 1, 2, 3$; so we must solve

$$\begin{aligned} \eta_1 + 2\eta_2 \qquad\quad &= 2 \\ 3\eta_1 + \eta_2 + \eta_3 &= 4 \\ 4\eta_3 &= 1 \end{aligned}$$

and this is a simple system of 3 equations in 3 unknowns whose unique solution is easily seen to be the one given in Sec. 2. Thus, knowing only the solution of the original problem we are able to find the solution of the dual.

Let us apply the equilibrium theorem to solving the transportation problem of Example 6. The proposed solution was given by the shipping matrix

0	0	4
2	3	1

We wish to find the prices π_1, π_2 and π'_1, π'_2, and π'_3. Note first that the supply σ_2 at P_2 is 7, but only 6 units are shipped out of P_2. According to the equilibrium theorem, therefore, the price $\pi_2 = 0$. Next, corresponding to the nonzero entries in the shipping matrix above we must have equations in the dual problem. These are

$$\begin{aligned} \pi_1' - \pi_2 &= \pi_1' = 2 \\ \pi_2' - \pi_2 &= \pi_2' = 4 \\ \pi_3' - \pi_1 &\quad\;\; = 3 \\ \pi_3' - \pi_2 &= \pi_3' = 6 \end{aligned}$$

so $\pi_1' = 2$, $\pi_2' = 4$, $\pi_3' = 6$, and $\pi_1 = 3$, which is the answer given in the previous section.

Bibliographical Notes

The term "linear programming" came into being about 1947–1948. The earliest published work containing these words in the title is due to Dantzig [1], *Programming* in a *Linear* Structure (italics mine), 1949. An early formulation and discussion of the diet problem was given by Stiegler [1] and of the transportation problem by Hitchcock [1]. The duality theorem was known to von Neumann [2] at least as early as 1947 and is contained in a set of privately circulated notes, but the proof as given there is incomplete. The first published proof based on von Neumann's notes is due to Gale, Kuhn, and Tucker [2]. The explicit statement of the equilibrium theorem is due to Goldman and Tucker [2] in a paper that contains a very complete treatment of linear programming theory.

Exercises

1. Prove that a transportation problem is feasible if and only if the total supply is at least equal to the total demand, that is,

$$\sum_{i=1}^{m} \sigma_i \geqq \sum_{j=1}^{n} \delta_j$$

2. Find a feasible shipping schedule for the transportation problem with 5 plants and 5 markets where the supplies and demands are given by

$$\begin{array}{llllll} & \sigma_1 = 120 & \sigma_2 = 75 & \sigma_3 = 205 & \sigma_4 = 145 & \sigma_5 = 90 \\ \text{and} & \delta_1 = 235 & \delta_2 = 50 & \delta_3 = 115 & \delta_4 = 80 & \delta_5 = 150 \end{array}$$

3. Show that the following standard maximum problem is not feasible: Find nonnegative numbers ξ_1 and ξ_2 which

$$\text{maximize } 3\xi_1 - 2\xi_2 \tag{1}$$

subject to

$$2\xi_1 + 5\xi_2 \leqq 3 \qquad (2)$$
$$-3\xi_1 + 8\xi_2 \leqq -5$$

4. Show that the following linear program is feasible but has no optimal solution: Find ξ_1, $\xi_2 \geqq 0$ such that

$$\xi_1 + \xi_2 \text{ is a maximum} \qquad (1)$$

subject to

$$-3\xi_1 + 2\xi_2 \leqq -1 \qquad (2)$$
$$\xi_1 - \xi_2 \leqq 2$$

5. Write the dual problem of the problem given in Exercise 4. In view of the result of Exercise 4 and the duality theorem what must be true of this dual problem? Verify this directly.

6. Construct a standard maximum problem involving two inequalities and two unknowns which has more than one optimal solution although not every feasible solution is optimal.

7. Consider the standard minimum problem of finding nonnegative numbers $\xi_1, \ldots, \xi_m$ which

$$\text{minimize} \sum_{i=1}^{m} \xi_i \gamma_i \qquad (1)$$

subject to

$$\sum_{i=1}^{m} \xi_i \alpha_{ij} \geqq \beta_j \qquad \text{for } j = 1, \ldots, n \qquad (2)$$

If $\alpha_{ij} \geqq 0$ for all i, j show that the problem is feasible if and only if

$$\beta_j \leqq 0 \qquad \text{whenever} \qquad \alpha_{1j} = \alpha_{2j} = \cdots = \alpha_{mj} = 0$$

Interpret this result for the case of the diet problem.

8. Write out the dual of the problem of Exercise 7. Assuming that the conditions of Exercise 7 hold, show from the duality theorem that both the original problem and its dual have optimal solutions if $\gamma_i \geqq 0$ for all i.

9. Verify that the dual of the diet problem is correctly described by the statement of Sec. 3, Subsec. 1.

10. Verify that the dual of the transportation problem is correctly given by the relations in Sec. 3, Subsec. 2.

11. Give an economic interpretation of the dual of the problem of "production to meet a given demand at minimum cost."

12. Consider the following standard maximum problem:

$$\text{maximize } \xi_1 + \xi_2 + \xi_3 + \xi_4 \tag{1}$$

subject to

$$\begin{aligned} \xi_1 + \xi_2 \qquad\qquad &\leqq 3 \\ \xi_3 + \xi_4 &\leqq 1 \\ \xi_2 + \xi_3 \qquad &\leqq 1 \\ \xi_1 \qquad + \xi_3 \qquad &\leqq 1 \\ \xi_3 + \xi_4 &\leqq 3 \end{aligned} \tag{2}$$

Show that this problem has the optimal solution

$$\xi_1 = 1 \qquad \xi_2 = 1 \qquad \xi_3 = 0 \qquad \xi_4 = 1$$

by finding a solution of the dual problem making use of the equilibrium theorem.

13. By the methods of Exercise 12 show that

$$\xi_1 = 4 \qquad \xi_2 = 1$$

is an optimal solution of the problem

$$\text{maximize } \xi_1 - \xi_2 \tag{1}$$

subject to

$$\begin{aligned} -2\xi_1 + \xi_2 &\leqq 2 \\ \xi_1 - 2\xi_2 &\leqq 2 \\ \xi_1 + \xi_2 &\leqq 5 \end{aligned} \tag{2}$$

14. The cost matrix for a transportation problem with 3 plants and 4 markets is

	M_1	M_2	M_3	M_4
P_1	4	4	9	3
P_2	3	5	8	8
P_3	2	6	5	7

The supplies and demands are

$$\sigma_1 = 3 \qquad \sigma_2 = 5 \qquad \sigma_3 = 7 \qquad \delta_1 = 2 \qquad \delta_2 = 5 \qquad \delta_3 = 4 \qquad \delta_5 = 4$$

We claim that the following shipping matrix gives an optimal solution:

0	0	0	3
0	5	0	0
2	0	4	1

Verify this by finding the prices in the dual solution, given the fact that the prices at P_2 and P_3 are zero.

15. Let $P_0, P_1, \ldots, P_n$ be a set of geographical points. A certain good is produced at P_0 and desired at P_n. For each pair of points P_i and P_j there is a nonnegative number c_{ij} called a *capacity*, which measures the maximum amount that can be shipped from P_i to P_j in one year. Formulate algebraically a linear program for maximizing the amount which can be received at P_n in a year. Show that the program is always feasible. (This is called the *maximum-flow problem.*) (Hint: Since goods are produced only at P_0 the flow out of P_i must not exceed the flow into P_i for $i > 0$.)

16. Write out the dual of the maximum-flow problem above. How many inequalities and unknowns does it contain? Show that this dual is always feasible if P_0 is distinct from P_n. In view of the duality theorem what does this imply for the original problem?

17. In a certain plant there are n different job openings $J_1, \ldots, J_n$ and m individuals $I_1, \ldots, I_m$ are available for working the various jobs. An efficiency expert has tested each individual at each job and has found that the *rating* of I_i for the job J_j is given by the nonnegative number α_{ij}. The problem is to determine what fraction of time I_i should work at job J_j assuming only one person can work at a given job at a time, in order to maximize the sum of the ratings. Formulate this problem algebraically as a linear program and show that it is always feasible. (This is called the *optimal-assignment* problem.)

18. Write out the dual of the optimal-assignment problem. Show that it has $m + n$ unknowns and mn inequalities. Using the duality theorem, prove that the optimal-assignment problem always has an optimal solution.

19. The following is the *rating matrix* for an optimal-assignment problem:

	J_1	J_2	J_3	J_4	J_5
I_1	12	9	10	3	8
I_2	6	6	2	2	9
I_3	6	8	10	11	9
I_4	6	3	4	1	1
I_5	11	1	10	9	12

We maintain that a solution to this problem is given by

I_1 works full time on J_1
I_2 works full time on J_5

I_3 works full time on J_4
I_4 works full time on J_2
I_5 works full time on J_3

Also, the 5 dual unknowns $\pi'_1, \ldots, \pi'_5$ corresponding to $J_1, \ldots, J_5$ in an optimal solution are

$$\pi'_1 = 3 \qquad \pi'_2 = 0 \qquad \pi'_3 = 1 \qquad \pi'_4 = 0 \qquad \pi'_5 = 3$$

Verify optimality by first finding the dual unknowns $\pi_1, \ldots, \pi_5$ corresponding to $I_1, \ldots, I_5$.