

# Mathematical in Games, Programming,

*Volume I*

MATRIX GAMES, PROGRAMMING,  
AND MATHEMATICAL ECONOMICS

# Methods and Theory and Economics

*by*

SAMUEL KARLIN

*Stanford University*



PERGAMON PRESS  
LONDON-PARIS  
1959

ADDISON-WESLEY PUBLISHING COMPANY, INC.  
READING, MASSACHUSETTS, U.S.A.  
LONDON, ENGLAND

*Copyright © 1959*

**ADDISON-WESLEY PUBLISHING COMPANY, INC.**

*Printed in the United States of America*

ALL RIGHTS RESERVED. THIS BOOK, OR PARTS THEREOF, MAY NOT BE REPRODUCED IN ANY FORM WITHOUT WRITTEN PERMISSION OF THE PUBLISHER.

*Library of Congress Catalog Card No. 60-5402*

---

Sole distributors in the United Kingdom:

PERGAMON PRESS LTD.

*4 and 5 Fitzroy Square, London W. 1*

PERGAMON PRESS S.A.R.L.

*24 Rue des Écoles, Paris V<sup>e</sup>*

## PREFACE

The mathematics involved in decision making is playing an increasingly important role in the analysis of management problems, in economic studies, in military tactics, and in operations research. These areas of application have led to challenging new types of mathematical structures. The development of game theory, linear and nonlinear programming, and mathematical economics marks one significant aspect of this kind of mathematics.

Essentially, these volumes are an attempt at a preliminary synthesis of the concepts of game theory and programming theory, together with the concepts and techniques of mathematical economics, into a single systematic theory. Secondarily, I hope they will prove useful as textbooks and reference books in these subjects, and as a basis for further study and research.

The coverage of these volumes is in no sense exhaustive; I have stressed most what appeals to me most. At the same time, I have tried to supply the necessary routine and basic information on which a proper understanding of game theory and programming depends. The presentation of each subject is designed to be self-contained and completely rigorous, but every effort has been made to bring out the conceptual depth and formal elegance of the over-all theory. In both volumes, the principles of game theory and programming are applied to a variety of simplified problems based on economic models, business decision, and military tactics in an attempt to clarify the key mathematical concepts involved and suggest their applicability to other similar problems.

Each chapter includes a certain amount of advanced exposition, which is generally incorporated in starred (\*) sections. Elementary background material is presented in the appendixes, and more advanced background material is incorporated directly in the text.

Each chapter contains problems of various degrees of difficulty, many of them leading to extensions of the theory. Solutions to most of the problems and hints for solving others are given at the end of each Part.

The three Parts (Parts I and II in Volume I, and Volume II, which comprises Part III) are independent of each other except for the material in starred sections, and may be studied independently in conjunction with the relevant appendixes. For the reader's convenience, a diagram showing the interrelationships of the various nonadvanced portions of Parts I and II of Volume I is presented on the following page.

Each chapter ends with a section of historical remarks, general comments on technical details, and citations of books and articles that may profitably be consulted for further information. All references cited are listed in the

bibliography at the close of each book. In some instances the historical remarks consist in assigning priorities to notable discoveries; any inaccuracies or omissions along these lines are wholly unintentional and deeply regretted.

My indebtedness extends to many, and not least to Stanford University, California Institute of Technology, the RAND Corporation, and the Office of Naval Research, all of which provided facilities, intellectual stimulation, and encouragement without which these books might never have been written.

Among my academic colleagues, I am grateful to Professors Bowker, Lieberman, Madow, Parzen, and Scarf of Stanford for their constant encouragement; to Professor H. F. Bohnenblust of California Institute of Technology, who introduced me to this field of study and who taught me more than I know how to acknowledge; to my students Rodriguez Restrepo and R. V. Miller, who wrote up my initial lectures on which the material of these books is based; to my students William Pruitt and Charles Stone, each of whom made valuable suggestions for organizing and improving the final manuscript; to Professors Hirofumi Uzawa and K. J. Arrow, who taught me most of what I know about mathematical economics; to my friends Melvin Dresher, Ray Fulkerson, and Harvey Wagner for their helpful comments on the first six chapters; and to Irving Glicksberg for assistance in developing the appendixes.

Finally, I am indebted to J. G. Bell, Jr., for helping to straighten out my syntax; to Frieda Gennerich, Sally Morgan, and Babette Doyle for their superb technical typing; and above all to my wife, who has offered warm encouragement, unbounded patience and kind consideration during these trying years of writing.

SAMUEL KARLIN

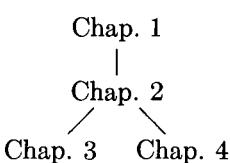
*Stanford, California*

*August 1959*

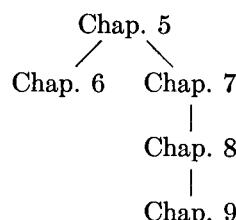
#### LOGICAL INTERDEPENDENCE OF THE VARIOUS CHAPTERS

*(applies only to unstarred sections of each chapter)*

##### PART I



##### PART II



## INTRODUCTION

### THE NATURE OF THE MATHEMATICAL THEORY OF DECISION PROCESSES

**Section 1. The background.** The art of making optimal judgments according to various criteria is as old as mankind; it is the essence of every field of endeavor from volleyball to logistics. The science of making such judgments, as opposed to the mere art, is a newer development, and the mathematical theory of decision-making is newer still—only a little over twenty years old. It has grown fast. Particularly in this last decade, we have seen a rapid spread of the use of scientific disciplines in economic planning, military strategy, business management, and systems engineering. There has been a steady interplay between the more careful formulation of business and logistics problems and the development of the mathematical tools needed for their solution. Along with and partly as a result of advances in game theory, linear programming, and queueing theory, there have grown up new practical disciplines whose very names were unknown forty years ago: operations research, management science, systems analysis, and a score of others. All these disciplines are concerned with the same fundamental task, that of scientifically analyzing various possible courses of action with a view to determining which course is "best" under the circumstances.

What a decision problem is must be clearly defined. Specifically, the possible courses of action, their consequences, the objectives of the various participants, and the nature of the random effects must be ascertained. From here on, the theory tries to evaluate and make comparisons of the various alternatives of action in terms of the objectives. A decision problem typically has four parts: (1) a model, expressing a set of assumed empirical relations among the set of variables; (2) a specified subset of decision variables, whose values are to be chosen by the firm or other decision-making entity; (3) an objective function of the variables, formulated in such a way that the higher its value, the better the situation is from the viewpoint of the firm; and (4) procedures for analyzing the effect on the objective function of alternative values of the decision variables.

The main component of a decision problem is the model. The following definition may help explain the significance and scope embodied in this concept. A model is a suitable abstraction of reality preserving the essential structure of the problem in such a way that its analysis affords insight into both the original concrete situation and other situations which have the same formal structure.

Model building has become increasingly widespread in these past years, particularly in the social sciences. Modern psychologists approach the study of learning processes in terms of a model involving stochastic variables. Econometricians, drawing on game theory and programming theory, have developed global equilibrium models in an attempt to understand the relationship of consumption, production, and price units in a competitive economy. Some political scientists are using the concepts of *n*-person game theory in an effort to analyze the power relationships that govern the decisions of legislative bodies. More and more business firms are applying rigorous mathematical thinking to management problems, inventory problems, production problems. Engineers are engaged in intensive studies in the new field of systems analysis, whose main technique consists of simulating idealized models of a given operation. The basic mathematics used in these studies is a mixture of the mathematical concepts and thinking processes of statistics, probability, and general decision theory.

The past fifty years have been years of fruitful fermentation in the disciplines of probability, stochastic processes, and statistics on the one hand, and have witnessed the development of game theory and linear programming on the other. These mathematical disciplines are the natural tools for the analytical investigation of problems arising in the social and managerial sciences. In this respect it is interesting to compare briefly the usefulness of probability, statistics, and decision theory with that of the other types of pure and applied mathematics.

The development of classical mathematics was principally inspired by problems based on phenomena in the fields of physics and engineering. In the usual classical physics problem, the variables of the system are seen to satisfy a system of well-defined deterministic relations. All parameters of the problem can be assumed known or accurately measurable, and one seeks to determine the relations between them in order to predict exactly what will happen in similar or analogous circumstances in the future. Until recently scientists were primarily concerned with just such deterministic problems, problems that could be solved, by and large, by the methods of partial differential equations and related mathematical techniques.

Clearly, however, these deterministic concepts and techniques are not altogether suitable for the social sciences, which typically present problems involving uncertainties and variability. Whenever human beings play a crucial role in the formulation of a model, we must allow for the unknown and sometimes random character of human actions, and for the randomness inherent in the environment in general, if we are to arrive at meaningful results. It is therefore inevitable that students of human behavior should be concerned prominently with the theories of probability, statistics, and decision analysis of the kind embodied in game theory and linear programming.

The reverse is also true: mathematicians and statisticians are increasingly concerned with the analytical problems originating from the analysis of human behavior. It is now no longer exclusively the physical sciences and engineering, but also the social and management sciences, that signal promising new directions for mathematical research.

There are two aspects to this research: the construction of the model, in which probability and stochastic processes play a fundamental role, and the normative essence of the model (optimization), the province of statistics, game theory, decision theory, and operations research. In practice these two aspects are inseparable. Obviously, to understand how to optimize the objective function, it is necessary to understand the structure and mechanism of the model; and to construct a model properly, it is necessary to understand what use is to be made of it.

In order to properly delimit the scope of these researches, we also mention the descriptive significance in the study of mathematical model building. This is in contrast to the normative essence of the model which we have emphasized. From the standpoint of mathematical models in psychology and sociology, the bulk of current literature is concerned with the orientation of the model toward empirical adequacy. On the other hand, the model based on situations in management, economics, and statistics is more concerned with the kind of normative questions of rational behavior theory (game and decision theory) to which we devote our main attention.

Inevitably the techniques of probability, statistics, and decision theory interact, and developments in each lead to developments in the others. As a result, the determination of optimum procedures for decision problems must be considered not in terms of half a dozen, or three, or two more or less separate sets of techniques, but in terms of a single over-all analytical methodology.

It is this methodology, in the context of game theory, programming, and mathematical economics, that we seek to delineate in this book. We may summarize our aims as an attempt to unify the mathematical techniques of the field and to help crystallize the concepts underlying these kinds of decision problems. It is not our purpose to propose a formal structure which will encompass all problems in the areas covered. We do not concentrate on definitions and abstractions *per se*. Rather, we attempt to develop in detail the principal mathematical methods that are applicable, putting emphasis wherever possible on mathematical prototypes and their analyses.

**Section 2. The classification of the mathematics of decision problems.** Our purpose in this section is to describe various useful mathematical classifications of decision problems that are also applicable to the theory of games, programming, and mathematical economics.

I. Decision problems can be divided into two contrasting types—static problems and dynamic problems. A static decision problem is one in which the variables do not involve time explicitly. A dynamic model is one in which time plays a very decisive role.

The distinction between static and dynamic problems is not always clear. Often a static situation is a stationary state of a dynamic situation, i.e., an equilibrium phase achieved by a dynamic process that has been in operation over a long period of time. Or an ostensibly dynamic process can be regarded as static, as when the same variables are introduced at the successive time points as new variables. Nevertheless there is an intrinsic difference in concept and technique between the two kinds of problems which will be manifest throughout the detailed analysis.

The principal technical difference between dynamic (multistage) models and static models is the degree of complexity involved in describing a given strategy or procedure. In a static situation a strategy is selected once and for all to be carried out directly, whereas the strategies available in the dynamic situation are usually complicated functions of information received and actions undertaken in the preceding stages. Theoretically, by defining a strategy suitably, we can reduce any dynamic model to a static model. In practice, however, this operation so complicates the nature of the strategy that it is often impossible to analyze the problem in the resulting terms. The essence of the analysis of a dynamic problem seems to be the exploitation of the recursive, time-dependent nature of the problem and the symmetries made available by the infinite horizon of time.

The significance of dynamic decision problems is obvious, since almost all human activity takes place in time. Events and learning processes are sequential. Most business decisions involve a time element. Economists try to take account of dynamic fluctuations in time. Nonetheless, the recognition and mathematical analysis of dynamic decision problems are of recent origin. A notable innovator in this regard was Wald, whose *Sequential Analysis of Statistical Data* [244]\* contained the essence of the notion of sequential decision-making for the accumulation of information. Other more special contributors to the sequential decision problem include Bellman [19], who saw the generality of the sequential idea and used it in the formulation of several multistage management decision models, and Arrow, Harris, and Marschak [8], who cast a type of inventory decision problem in dynamic form. The theory of multistage games is an extension of the usual theory of games to a dynamic situation. A recent volume [72] shows the importance of this development.

The study of static models serves two purposes. First, a static model may be the best way of describing a static problem whose solution is inter-

---

\* References are to the Bibliography.

esting in itself and affords insight into the workings and structure of a given operation. One common static problem of this sort is that of optimizing a given criterion among different procedures where the variables range over a prescribed set in which the time variable is not a special variable. Second, from a technical point of view, many dynamic models must be approached through the study of modified static problems. For example, it is often convenient to reduce the infinite version of the dynamic problem to the case of a finite number of periods  $N$ . The solution of the resulting problem is obtained by a simple extension of the static model, after which a passage to the limit on  $N$  produces the solution of the full dynamic problem.

II. Decision problems can also be classified in terms of the number of antagonistic participants involved. We note three examples.

(1) If there are at least two irreconcilable participants, we are dealing with problems in the theory of games. Where there are just two such opponents, their objectives are clear: each strives to achieve as great a return as possible, knowing that his opponent will strive for the same objective with all his ability and intelligence. Where there are more than two, their objectives are not so clear. The conceptual framework of this model has not yet been well established, and therefore it is difficult to develop appropriate mathematical methodology for it.

(2) Where there is only one participant involved and his objectives are clearly defined, the basic decision problem reduces technically to a strict maximization of the objective functions subject to the natural constraints of the model. Generally, this situation leads to variational problems, with the solution usually occurring on the boundaries of the domain of the variables. To solve such extreme constrained problems entails developing new techniques. The fields of linear and dynamic programming abound with models of this kind.

(3) Where there is one participant and uncertainties are present, a solution can usually be obtained by combining the methods of statistics with variational techniques.

III. A third breakdown of the general decision model is into deterministic and stochastic cases. Naturally the deterministic models are conceptually and technically much simpler and more frequently amenable to thorough analytical treatment. However, the majority of practical problems involve uncertainties, variability, and all kinds of randomness. To solve problems of this sort requires a thorough acquaintance with the theory of stochastic processes and the corresponding theory of inference in stochastic processes.

IV. Another classification can be made in terms of the complexity of the available strategies (the strategy space). Strategy spaces can be divided into two categories: finite-dimensional and infinite-dimensional. In many

of our examples, the infinite-dimensional space is identifiable with a subset of some classical function space. When this is so, the specific structure of this function space may be employed in seeking a solution. This method is used to solve so-called games played over the unit square; it is also useful in the more general finite-action statistical decision problems.

V. Finally, a classification can be made in terms of subject matter: i.e., into game theory, linear programming, multistage decision problems, statistical decision theory, etc. Although many of these subject areas overlap both conceptually and technically, each has certain predominant features which can be used to solve problems that are not amenable to solution by the others.

It is impossible to treat all kinds of decision problems adequately in a single work. We have accordingly devoted these two volumes to problems of game theory, programming, and mathematical economics, leaving some aspects of statistical decision theory and a discussion of dynamic stochastic decision problems to other books now in the making.

Broadly speaking, then, the present two volumes deal primarily with static decision problems in which uncertainties and randomness in the structure of the model are not central. We devote Part I (four chapters) to finite game theory, and Part II (five chapters) to linear programming, nonlinear programming, and mathematical economics. The affinity of mathematical economics to programming methods, concepts, and applications justifies this grouping. In the main these two parts treat static, finite decision problems. The concept of mixed strategy for finite games (implying an element of randomness) comes into the solution of game problems but does not appear in the original formulation of game theory.

Volume II (nine chapters) is concerned with infinite static game theory, in the sense that the strategy spaces are infinite-dimensional and the game has the basic form of involving one play.

**Section 3. The main disciplines.** In this section, by way of indicating how game theory and programming theory fit into the larger picture, we present a cursory description of the four main disciplines of decision theory. It should be understood, of course, that proper insight into any of these disciplines can be gained only from a detailed study of specific problems and solutions.

1. *Theory of games.* Throughout these volumes we are concerned primarily with two-person, zero-sum games. The mathematical investigations done on the theory of games involving more than two participants is of an introductory nature, primarily because of the conceptual difficulties inherent to the formulation of the model. There has been some elegant and sophisticated work on "solutions" of  $n$ -person game models [215, 169], but the techniques and concepts of the subject are relatively undeveloped

and we have therefore decided to omit them from consideration in this book. The challenge of  $n$ -person game theory is being taken up, however, and we may hope that the conceptual obstacles will soon be overcome.

In two-person, zero-sum games, each player has available a space of strategies, and there is a known payoff function which represents the yield to Player I as a function of the specific strategies employed by both players. Of course, the yield to Player II is the negative of Player I's yield.

Player I seeks to maximize his return, but affects the pay-off only through control of his own strategy; Player II desires to minimize Player I's yield, since the interests of the players are in strict conflict. This leads to the min-max principle. It is remarkable that one can define a rational mode of behavior for each of two players with irreconcilable opposing interests. The proof of this fact was the fundamental contribution of John von Neumann.

The strict conflict of interests is a realistic assumption for parlor games and military operations, and for a few special situations in business, notably duopoly. Outside of this rather limited domain, one is hard pressed to come up with a truly realistic game problem. This is of little importance, however; the fundamental contribution of game theory is not its solution of practical problems, but its clarification of the basic methods of decision theory. Advances and refinements in game theory have fundamentally influenced the development of statistics, linear and nonlinear programming, mathematical economics, and model building in general.

2. *The theory of programming.* The theory of linear and nonlinear programming is simply the mathematics of maximizing a function of several variables which are subject to constraints. As the adjective "linear" suggests, the constraints and the function maximized are often linear functions of the variables. Unfortunately, the ordinary calculus is usually not applicable to programming problems, since their solutions necessarily fall on the boundaries of the domain over which the free variables range. It is accordingly the task of the theory to devise analytical methods of determining solutions wherever possible, and in the absence of such methods to furnish efficient computational means of approximating a solution. Moderate success has been achieved in both directions.

On the theoretical side, the theory of linear programming is equivalent to the theory of discrete two-person, zero-sum games. The interaction between these two theories has benefited and accelerated the growth of both. An extension of the theory to deal with nonlinear functions has also been partly accomplished.

The applications of linear programming are numerous, especially in the field of activity analysis, which is concerned with planning production operations. A typical linear programming model is the assignment problem, which calls for the assignment of  $n$  persons of various talents

to  $n$  jobs of varying importance in such a way as to maximize the total value of the work done in those jobs. So general a model obviously has any number of practical applications, and the same is true of several other key programming models. The oil industry has particularly benefited from the use of linear programming to determine efficient production procedures.

3. *Multistage decision problems.* Multistage decision problems, sometimes called dynamic programming, are problems involving a number (usually an infinite number) of decisions made in the course of time. Each decision depends on all past decisions and in turn affects all future decisions. A typical example is furnished by the inventory problem. To what level should a merchant raise his available stock in order to maximize his gain? The factors influencing his decision are ordering costs, profits, holding costs, and penalties (e.g., loss of trade, loss of good will) for not having goods available to satisfy demand. A further complication could arise from a lag in delivery of the order he places. This ordering problem comes up periodically, which makes the problem dynamic. Of course what happened in preceding periods will inevitably affect the merchant's decisions.

The same general considerations apply to planning production schedules, determining replacement policies for old equipment, releasing hydroelectric energy, and the like. Basic dynamic models are found also in the fields of sociology and psychology, notably in the study of learning processes and in latent structure analysis. Queueing models furnish almost an inexhaustible source of dynamic decision processes, with important applications to cargo loading, transportation systems, doctor's schedules, etc.

Finally, statistical sequential analysis obviously belongs to the category of multistage procedures. In fact, this is the richest subject of all from the point of view of mathematical content, subtlety, and depth, since it draws on some of the finer points of the theory of stochastic processes, statistical decision theory, game theory, and general decision methodology.

4. *Statistical decision theory.* Statistical decision theory as conceived by Wald is essentially statistics considered from a broader point of view. Almost all types of statistical analysis—testing hypotheses, estimating unknown parameters, prediction, discrimination—can be cast into this general framework, and their relevance and effectiveness can be evaluated in its terms. The spirit and mathematical procedures of statistical decision theory and the theory of games are very much the same, but their objectives differ. Explicitly, the theory of games deals primarily with two conflicting interests, whereas statistical decision theory is concerned with determining optimal procedures, with allowances for uncertainties in nature. In a sense the statistician and nature represent two players engaged in a game, but the relationship between them is not the irreconcilable antagonism upon which the theory of games is predicated.

The general outlook of statistical decision theory has helped increase our understanding of many of the standard principles utilized in statistical practice: for example, the principle of unbiasedness, type-A regions, invariance concepts, and min-max theory, which all represent criteria by which we single out a given mode of action. The classical Neyman-Pearson philosophy of testing hypotheses and estimating unknown parameters by means of confidence intervals is encompassed within the boundaries of statistical decision theory. The scope of this theory is clearly tremendous, and its future influence on the development of specific statistical theory and practice should be very great.

## NOTATION

### VECTORS

A vector  $\mathbf{x}$  with  $n$  components is denoted by

$$\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle.$$

$\mathbf{x} \geq \mathbf{0}$  shall mean  $x_i \geq 0$  ( $i = 1, 2, \dots, n$ ).

$\mathbf{x} > \mathbf{0}$  shall signify  $x_i \geq 0$  and at least one component of  $\mathbf{x}$  strictly positive.

$\mathbf{x} \gg \mathbf{0}$  denotes that  $x_i > 0$  ( $i = 1, 2, \dots, n$ ) (all components are positive).

The inner product of two real vectors  $\mathbf{x}$  and  $\mathbf{y}$  in  $E^n$  is denoted by

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i.$$

In the complex case

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \bar{y}_i,$$

where  $\bar{y}_i$  denotes conjugate complex.

The distance between two vectors  $\mathbf{x}$  and  $\mathbf{y}$  is denoted by  $|x - y|$ .

### MATRICES

A matrix  $\mathbf{A}$  in terms of its components is denoted by  $\|a_{ij}\|$ .

A matrix  $\mathbf{A}$  applied to vector  $\mathbf{x}$  in the manner  $\mathbf{x}\mathbf{A}$  gives the vector

$$\langle \sum x_i a_{i1}, \sum x_i a_{i2}, \dots, \sum x_i a_{in} \rangle.$$

Similarly,

$$\mathbf{Ay} = \langle \sum a_{1j} y_j, \sum a_{2j} y_j, \dots, \sum a_{nj} y_j \rangle.$$

$(\mathbf{Ax})_j$  denotes the  $j$ th component of the vector  $\mathbf{Ax}$ .

The transpose of  $\mathbf{A}$  is denoted by  $\mathbf{A}'$ .

The determinant of  $\mathbf{A}$  is usually designated by  $|\mathbf{A}|$  and alternatively by  $\det \mathbf{A}$ .

## CONVEX SETS

The convex set spanned by a set  $S$  is denoted alternatively by  $\text{Co}(S) = [S]$ .

The convex cone, spanned by  $S$ , is denoted by  $P_S$  or  $\mathcal{P}_S$ .

## DISTRIBUTIONS

The symbols

$$\mathbf{x}_{t_0}, \mathbf{y}_{t_0} \quad \text{and} \quad \mathbf{I}_{t_0} \quad (0 \leq t_0 \leq 1)$$

are used interchangeably to represent a distribution defined on the unit interval which concentrates its full mass at the point  $t_0$ , that is,

$$\mathbf{x}_{t_0}(\xi) = \begin{cases} 0 & \xi < t_0, \\ 1 & \xi \geq t_0. \end{cases}$$

The symbol

$$\sum_{i=1}^k \lambda_i I_{\xi_i}$$

represents the probability distribution function with jumps  $\lambda_i$  located at  $\xi_i$ .

# CHAPTER 1

## THE DEFINITION OF A GAME AND THE MIN-MAX THEOREM

**1.1 Introduction. Games in normal form.** The mathematical theory of games of strategy deals with situations involving two or more participants with conflicting interests. The outcome of such games is usually controlled partly by one side and partly by the opposing side or sides; it depends to some extent on chance, but primarily on the intelligence and skill employed by the participants. Aside from games proper, such as poker and chess, there are many conflicting situations to which the theory of games can be applied, notably in certain areas of operations research, economics, politics, and military science.

We shall first consider only two-person zero-sum games, i.e., games with only two participants (competing persons, teams, firms, nations) in which one participant wins what the other loses. It should be noted that the terms of this situation exclude the possibility of any bargaining between the participants. We shall also be concerned with two-person constant-sum games, in which the two players compete irreconcilably for the greatest possible share of the kitty. By suitable renormalization such a game may be converted into a zero-sum game.

A fundamental concept in game theory is that of *strategy*. A strategy for Player I is a complete enumeration of all actions Player I will take for every contingency that might arise, whether the contingency be one of chance or one created by a move of the opposing player. What a strategy is should not be interpreted too naively. It might seem that once Player I has chosen a strategy, every move he makes at any stage is determined, in the sense that he has in mind at the beginning of the game a sequence of moves which he will carry out no matter what his opponent does. However, what we mean by "strategy" is a rule which in determining Player I's  $i$ th move takes into account everything that has happened before his  $i$ th turn. The reason a player does not change a strategy during a game is not that the strategy has committed him to a sequence of moves he must make no matter what his opponent does, but that it gives him a move to make in any circumstances that may arise.

It is usual, in describing a game, to regard all possible procedures, good or bad, as possible strategies. Even in simple games the number of possible

strategies is often forbidding. Consider the game of ticktacktoe.\* Suppose Player I makes the first move. There are nine possible positions for his first cross. Player II then has eight possible moves he can make, and what Player I does at his second opportunity to move will depend on the preceding move of Player II. In this situation there are seven possible moves he may make. Player I's third move will, of course, depend on all the preceding moves of both players; and so on.

A strategy might start off as follows: Player I's first cross is to be made in the upper right-hand square. If Player II marks either the square below or the square to the left, Player I makes his second cross in the center square; if Player II marks the center square, Player I makes his second cross in the lower left-hand square; if Player II marks any of the other five squares, Player I makes his second cross in the square immediately below the first; and thus the description goes on. Player I might even embody in his strategy the possibility of randomizing, according to a fixed probability distribution, among alternatives at a given move.

Clearly, an enormous number of possible strategies present themselves even for such a simple game as ticktacktoe. Although many of them intuitively seem to be poor strategies, we are obliged to include all the possibilities in order to give a complete description of the game. In the course of this book mathematical tools for manipulating and analyzing these large sets of strategies will be developed.

A second fundamental concept in game theory is that of the *pay-off*. The pay-off is the connecting link between the set of strategies open to Player I and the set open to Player II. Specifically, it is a rule that tells how much Player I may be expected to win from Player II if Player I chooses any particular strategy from his set of strategies and Player II chooses any particular strategy from his set. The pay-off function is always evaluated in terms of the appropriate utility units (see the notes to this section on p. 34).

We are now ready for a formal definition of a game.

A game is defined to be a *triplet*  $\{X, Y, K\}$ , where  $X$  denotes the space of strategies for Player I,  $Y$  signifies the space of strategies of Player II, and  $K$  is a real-valued function of  $X$  and  $Y$ . Player I chooses a strategy  $\mathbf{x}$  from  $X$  and Player II chooses a strategy  $\mathbf{y}$  from  $Y$ . For the pair  $\{\mathbf{x}, \mathbf{y}\}$  the pay-off to Player I is  $K(\mathbf{x}, \mathbf{y})$  and the pay-off to Player II is  $-K(\mathbf{x}, \mathbf{y})$ . We shall call  $K$  the pay-off kernel.

In the absence of a statement to the contrary, the following conditions are assumed to be satisfied throughout this chapter:

---

\* Ticktacktoe is played on a  $3 \times 3$  matrix grid. Players move alternately and on each turn are allowed to capture one of the remaining free squares. The first player who takes possession of three squares which are on a single horizontal, vertical, or diagonal line wins.

- (a)  $X$  is a convex, closed, bounded set in Euclidean  $n$ -space  $E^n$ .
- (b)  $Y$  is a convex, closed, bounded set in Euclidean  $m$ -space  $E^m$ .
- (c) The pay-off kernel  $K$  is a convex linear function of each variable separately. Explicitly,

$$K[\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2, \mathbf{y}] = \lambda K(\mathbf{x}_1, \mathbf{y}) + (1 - \lambda) K(\mathbf{x}_2, \mathbf{y})$$

and

$$K[\mathbf{x}, \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2] = \lambda K(\mathbf{x}, \mathbf{y}_1) + (1 - \lambda) K(\mathbf{x}, \mathbf{y}_2),$$

where  $\lambda$  is a real number satisfying  $0 \leq \lambda \leq 1$ .

Several of these limitations will be relaxed in later chapters.

The property that is essential in Chapters 1–4 is that  $X$  and  $Y$  are convex sets and have the character of finite dimensionality. (A representation of a game as a triplet involving strategy spaces which are finite-dimensional is necessarily a restriction; its justification rests on the fact that numerous actual games are of this kind.) The identification of strategies with points in Euclidean  $n$ -space is a convenience that simplifies the mathematical analysis.

An important special class of games is obtained where  $X$  is taken as the simplex  $S^n$  in  $E^n$ , defined as the set of all  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  where  $x_i \geq 0$  and

$$\sum_{i=1}^n x_i = 1,$$

and the space  $Y$  is the corresponding simplex  $T^m$  in  $E^m$ . The pay-off kernel then takes the form

$$K(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^m \sum_{i=1}^n x_i a_{ij} y_j = (\mathbf{x}, \mathbf{Ay}),$$

where  $\mathbf{A}$  is the matrix  $\|a_{ij}\|$ . In the case of such matrix games we shall often denote the pay-off corresponding to strategies  $\mathbf{x}$  and  $\mathbf{y}$  as  $A(\mathbf{x}, \mathbf{y})$  in place of  $K(\mathbf{x}, \mathbf{y})$  to suggest that these games are matrix games, i.e., games in which  $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{Ay})$ .

Certain special strategies consisting of vertex points of  $X$  are denoted by  $\boldsymbol{\alpha}_i = \langle 0, \dots, 0, 1, 0, \dots, 0 \rangle$  ( $i = 1, \dots, n$ ), where the 1 occurs in the  $i$ th component. These are Player I's *pure strategies*. Similarly, the strategies  $\boldsymbol{\beta}_j = \langle 0, \dots, 0, 1, 0, \dots, 0 \rangle$  of  $Y$  ( $j = 1, \dots, m$ ) are referred to as Player II's pure strategies. Since  $A(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = K(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = a_{ij}$ , we see that the  $i, j$  element of the matrix array  $A$  expresses the yield to Player I when Player I uses the pure strategy  $\boldsymbol{\alpha}_i$  and Player II employs the pure strategy  $\boldsymbol{\beta}_j$ .

A strategy  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  with no component equal to 1 is called a *mixed strategy*. In view of the relationship

$$\sum_{i=1}^n x_i K(\alpha_i, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}), \quad (1.1.1)^*$$

the mixed strategy  $\mathbf{x}$  can be effected as follows. An experiment is conducted with  $n$  possible outcomes such that the probability of the  $i$ th outcome is  $x_i$ . The  $i$ th pure strategy is used by Player I if and only if the  $i$ th outcome has resulted, and the yield to Player I becomes the expected yield from this experiment, namely

$$\sum_{i=1}^n x_i K(\alpha_i, \mathbf{y}).$$

This amounts to playing each pure strategy with a specified probability. We can therefore interpret a mixed strategy as a probability distribution defined on the space of pure strategies, and conversely. Similar interpretations apply to the strategy space  $T^m$  of Player II.

The usual introductory approach to game theory is to enumerate the spaces of pure strategies for both players and to specify the pay-off matrix  $\mathbf{A}$  corresponding to these pure strategies. The concept of mixed strategy is introduced subsequently, and the pay-off function is then replaced in a natural way by the expected pay-off function. In this book we have started with the more general formulation of a game, in terms of a triplet specifying the complete strategy spaces for both players and the pay-off kernel. In this formulation, the distinction between pure and mixed strategies does not exist. Nevertheless, in many special classes of games it is natural to single out the pure strategies, which span convexly the space of all strategies. This will be done wherever it is advantageous, e.g., whenever pure strategies occur in a natural manner and possess special significance. Broadly speaking, however, we lose no generality by starting directly with two-person zero-sum games in normal form, where the strategy spaces are finite-dimensional.

Again, it has been common in the literature of game theory to make a distinction between games in extensive form and games in normal form and to take the first as a point of departure. A game formulated in extensive form is developed in terms of more primitive concepts such as "play," "chance move," "personal move," and "information structure." A strategy

---

\* Although we have postulated this formula, it is possible to derive it by appealing to a suitable axiom system satisfied by a preference pattern which selects among alternative probability distributions over the space of outcomes resulting from the choice of a pure strategy (see the notes to this section).

is then defined within the framework of these notions, and the analysis of optimal strategies proceeds from this point. Finally, a theorem is proved to the effect that any game in extensive form may be in fact reduced to an equivalent game in normal form.

In contrast, our definition of a game as a triplet begins immediately with the concepts of strategy and pay-off, and is flexible and general enough to encompass all forms of finite game theory, including in particular the structure of games in extensive form. By carefully defining strategies and specifying completely  $X$ ,  $Y$  and  $K(\mathbf{x}, \mathbf{y})$  we are able to handle all forms of information patterns that arise. This will become clear as we study specific games.

It may happen that in special instances one can construct two apparent strategy spaces which are in fact equivalent in terms of pay-off. Whenever necessary we shall demonstrate this equivalence. For purposes of mathematical consistency, however, whenever any two given games differ in a specific component, the two games are taken to be distinct. For example, if we enlarge the  $X$  strategy space while the components  $Y$  and  $K$  remain unchanged, we create a new game. This is so even if the added strategies are obviously inferior and cannot affect either player's ultimate choice of an optimal strategy. In practice, as will be seen, the strategy spaces exhibited constitute an exhaustive class of procedures, i.e., a class of procedures that take into account all the fine structure of the model.

**1.2 Examples.** In dealing with finite matrix games it is sufficient to specify the pure strategies for both players and the corresponding pay-off matrix  $\|a_{ij}\|$ . The pay-off kernel for arbitrary mixed strategies  $\mathbf{x}$  and  $\mathbf{y}$  is given by the expression

$$K(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^m \sum_{i=1}^n x_i a_{ij} y_j. \quad (1.2.1)$$

*Example 1. Matching pennies.* Players I and II each display simultaneously a single penny. If I matches II, i.e., if both are heads or both are tails, I takes II's penny. Otherwise, II takes I's penny. The pay-off kernel is represented in matrix form as follows:

$$\begin{array}{c} & \text{Player II} \\ & \text{H} \quad \text{T} \\ \text{Player I} & \begin{array}{c|cc} \text{H} & 1 & -1 \\ \text{T} & -1 & 1 \end{array} \end{array}$$

The first pure strategy for I would be to display heads and the second

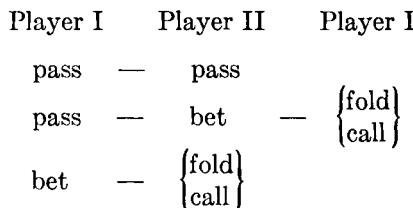
tails. One possible mixed strategy would be to randomize equally between showing heads and tails (i.e.,  $\mathbf{x} = \langle \frac{1}{2}, \frac{1}{2} \rangle$ ).

*Example 2. Two-finger Morra.* Each player displays either one or two fingers and simultaneously guesses how many the opposing player will show. If both players guess correctly or both guess incorrectly, the game is a draw. If only one guesses correctly, he wins an amount equal to the total number of fingers shown by both players.

In this case each pure strategy will have two components: (a) the number of fingers to show, and (b) the number of fingers to guess. Thus, each strategy can be represented by a pair  $\langle a, b \rangle$ , where  $a$  denotes the first component and  $b$  the second. For example, the strategy  $\langle 2, 1 \rangle$  for Player I is to show two fingers and guess one. There will be four such pure strategies for each player:  $\langle 1, 1 \rangle$ ,  $\langle 1, 2 \rangle$ ,  $\langle 2, 1 \rangle$ , and  $\langle 2, 2 \rangle$ . The pay-off matrix is shown below.

		Player II			
		$\langle 1, 1 \rangle$	$\langle 1, 2 \rangle$	$\langle 2, 1 \rangle$	$\langle 2, 2 \rangle$
Player I	$\langle 1, 1 \rangle$	0	2	-3	0
	$\langle 1, 2 \rangle$	-2	0	0	3
	$\langle 2, 1 \rangle$	3	0	0	-4
	$\langle 2, 2 \rangle$	0	-3	4	0

*Example 3. Poker model.* In this poker model there are only three possible hands, as compared with  $\binom{52}{5}$  in regular poker, and all three are considered equally likely to occur. One hand is dealt to each player. There is a preference ordering among the hands: Hand 1 wins over hands 2 and 3, and hand 2 wins over hand 3. The ante is  $a$  units. Player I can either pass or bet  $b$  units. If Player I bets then Player II can either fold or call. If Player I passes and Player II bets, then Player I again has the option of folding or calling. The three possible courses of play can be diagrammed as follows:



If a pass follows a bet, the bet wins. If both contestants pass or if one contestant calls, the hands are compared and the player with the better hand wins the pot.

A strategy for Player I must provide a set of actions to be taken for each of the three possible hands that he might draw and for all possible contingencies of the play of a hand. We can denote such a strategy by a triplet of pairs  $\{\{\alpha, \beta\}_1, \{\alpha, \beta\}_2, \{\alpha, \beta\}_3\}$ . The component  $\alpha$  of  $\{\alpha, \beta\}_k$  represents for Player I the probability of betting on the first round if he draws hand  $k$ . The component  $\beta$  represents the probability of his betting on the second round if he holds hand  $k$ . The second component is relevant only if there is a second round, but since a strategy must take all contingencies into account, a strategy in this game must include the values of  $\beta$ .

Similarly, Player II's strategies can be represented by a triplet of pairs  $\{\{\lambda, \mu\}_1, \{\lambda, \mu\}_2, \{\lambda, \mu\}_3\}$ . The component  $\lambda$  of  $\{\lambda, \mu\}_k$  is the probability of Player II's contesting a bet by Player I if Player II has drawn hand  $k$ . The component  $\mu$  represents the probability of Player II's betting if Player I has previously passed.

In this example the mixed strategies have been introduced directly. In view of the fact that the components of such a strategy are formulated in terms of the behavior at each move of both players, strategies of this sort are often called "behavior strategies." This specific example is fully discussed in Section 4.3. For other examples see Problems 1 through 4 at the end of this chapter.

**1.3 Choice of strategies.** Suppose that Player II is compelled to announce to Player I what strategy he is going to use, and that he announces  $\mathbf{y}_0$ . Then Player I, seeking to maximize his pay-off or yield, naturally chooses his strategy  $\mathbf{x}_0$  so that  $K(\mathbf{x}_0, \mathbf{y}_0) = \max_{\mathbf{x}} K(\mathbf{x}, \mathbf{y}_0)$ . The best thing for Player II to do under these circumstances would be to announce  $\mathbf{y}_0$  such that

$$\max_{\mathbf{x}} K(\mathbf{x}, \mathbf{y}_0) = \min_{\mathbf{y}} \max_{\mathbf{x}} K(\mathbf{x}, \mathbf{y}) = \bar{v},$$

where  $\bar{v}$  (the upper value) can be interpreted as the most that Player I can achieve if Player II employs the strategy  $\mathbf{y}_0$ .

Suppose the tables are turned and Player I has to announce his strategy  $\mathbf{x}_0$ . Since Player II is sure to choose  $\mathbf{y}_0$  such that

$$K(\mathbf{x}_0, \mathbf{y}_0) = \min_{\mathbf{y}} K(\mathbf{x}_0, \mathbf{y}),$$

Player I can best protect himself by choosing  $\mathbf{x}_0$  such that

$$\min_{\mathbf{y}} K(\mathbf{x}_0, \mathbf{y}) = \max_{\mathbf{x}} \min_{\mathbf{y}} K(\mathbf{x}, \mathbf{y}) = \underline{v},$$

where  $\underline{v}$  (the lower value) can be interpreted as the most that Player I can guarantee himself independent of Player II's choice of strategy. In order to continue this line of reasoning, we first establish the following simple result.

► LEMMA 1.3.1. (a) Let  $f(\mathbf{x}, \mathbf{y})$  denote a real-valued function defined on  $X \times Y$ ; then

$$\inf_y \sup_x f(\mathbf{x}, \mathbf{y}) \geq \sup_x \inf_y f(\mathbf{x}, \mathbf{y}) \quad (\pm \infty \text{ are possible values}). \quad (1.3.1)$$

(b) If  $X$  and  $Y$  are compact and  $f$  is continuous, then sup and inf in (1.3.1) may be replaced by max and min, respectively.

*Proof.* By definition, for any  $\mathbf{y}$ ,  $\sup_x f(\mathbf{x}, \mathbf{y}) \geq f(\mathbf{x}, \mathbf{y})$ . Hence  $\inf_y \sup_x f(\mathbf{x}, \mathbf{y}) \geq \inf_y f(\mathbf{x}, \mathbf{y})$ , and (1.3.1) follows immediately. The simple proof of (b) is omitted.

Lemma 1.3.1 for  $f = K$  gives

$$\min_{\mathbf{y} \in Y} \max_{\mathbf{x} \in X} K(\mathbf{x}, \mathbf{y}) \geq \max_{\mathbf{x} \in X} \min_{\mathbf{y} \in Y} K(\mathbf{x}, \mathbf{y}).$$

Expressed in terms of the upper and lower values of a game, we obtain

$$\bar{v} \geq \underline{v}.$$

Our principal aim in this chapter is to establish that for a matrix game  $\bar{v} = \underline{v} = v$ . By an appropriate choice of strategy Player I can guarantee himself the value  $\underline{v} = v$ , and by judicious play Player II can prevent Player I from achieving more than  $\bar{v} = \underline{v}$ . Thus, unless Player I has additional information about Player II's mode of behavior—for example, expects him to choose a strategy in a wild manner—Player I should play so as to achieve  $v$ . If he departs from the course of action that assures him the value  $v$ , his ultimate yield might be less than  $v$ . Thus, it is reasonable to call this common value  $v$  the value of the game to Player I. Of course,  $-v$  is the value to Player II.

All strategies  $\mathbf{x}_0$  and  $\mathbf{y}_0$  such that  $K(\mathbf{x}_0, \mathbf{y}) \geq v$  for all  $\mathbf{y}$  and  $K(\mathbf{x}, \mathbf{y}_0) \leq v$  for all  $\mathbf{x}$  will be referred to as optimal strategies for Players I and II, respectively.

The following simple criterion is often very useful in determining when  $\bar{v} = \underline{v}$ ; moreover, it points up the connection between the existence of optimal strategies and the equality of the lower and upper values.

► THEOREM 1.3.1. If there exist  $\mathbf{x}_0 \in X$ ,  $\mathbf{y}_0 \in Y$  and a real number  $v$  such that

$$K(\mathbf{x}_0, \mathbf{y}) \geq v \quad \text{for all } \mathbf{y} \in Y$$

and

$$K(\mathbf{x}, \mathbf{y}_0) \leq v \quad \text{for all } \mathbf{x} \in X,$$

then

$$\bar{v} = \min_y \max_x K(\mathbf{x}, \mathbf{y}) = v = \max_x \min_y K(\mathbf{x}, \mathbf{y}) = \underline{v},$$

and conversely.

*Proof.* (a) Since

$$K(\mathbf{x}_0, \mathbf{y}) \geq v \quad \text{for all } \mathbf{y},$$

it follows that

$$\min_y K(\mathbf{x}_0, \mathbf{y}) \geq v \quad \text{and} \quad \max_x \min_y K(\mathbf{x}, \mathbf{y}) \geq v.$$

Similarly,

$$\max_x K(\mathbf{x}, \mathbf{y}_0) \leq v;$$

hence

$$\min_y \max_x K(\mathbf{x}, \mathbf{y}) \leq v.$$

This gives

$$\max_x \min_y K(\mathbf{x}, \mathbf{y}) \geq \min_y \max_x K(\mathbf{x}, \mathbf{y}).$$

But

$$\min_y \max_x K(\mathbf{x}, \mathbf{y}) \geq \max_x \min_y K(\mathbf{x}, \mathbf{y})$$

by Lemma 1.3.1. Therefore

$$\min_y \max_x K(\mathbf{x}, \mathbf{y}) = \max_x \min_y K(\mathbf{x}, \mathbf{y}).$$

(b) Conversely, choose  $\mathbf{y}_0$  such that

$$\max_x K(\mathbf{x}, \mathbf{y}_0) = \min_y \max_x K(\mathbf{x}, \mathbf{y}) = v,$$

and  $\mathbf{x}_0$  such that

$$\min_y K(\mathbf{x}_0, \mathbf{y}) = \max_x \min_y K(\mathbf{x}, \mathbf{y}) = v.$$

Then  $K(\mathbf{x}, \mathbf{y}_0) \leq v$  and  $K(\mathbf{x}_0, \mathbf{y}) \geq v$ . This completes the proof.

► **COROLLARY 1.3.1.** A necessary and sufficient condition that

$$\min_y \max_x K(\mathbf{x}, \mathbf{y}) = \max_x \min_y K(\mathbf{x}, \mathbf{y})$$

is that there exist strategies  $\{\mathbf{x}_0, \mathbf{y}_0\}$  such that for all  $\mathbf{y}$  in  $Y$  and all  $\mathbf{x}$  in  $X$

$$K(\mathbf{x}_0, \mathbf{y}) \geq K(\mathbf{x}_0, \mathbf{y}_0) \geq K(\mathbf{x}, \mathbf{y}_0).$$

Obviously,  $v = K(\mathbf{x}_0, \mathbf{y}_0)$ .

**1.4 The min-max theorem for finite matrix games.** Throughout this section we shall be concerned exclusively with matrix games. For this special case the strategy spaces  $X$  and  $Y$  of Players I and II are identifiable with the simplexes  $S^n$  and  $T^m$ , respectively, and thus each set of all strategies is spanned convexly by the pure strategies. Our objective is to establish the fundamental min-max theorem for these matrix games.

The number of different proofs of this theorem is very large. Von Neumann constructed two proofs: one based on the Brouwer fixed-point theorem (see Section C.2 of the Appendix), the other on the proposition that nonoverlapping convex sets lying in a Euclidean space can be separated by hyperplanes. Other proofs have been given which involve algebraic inequalities and the properties of convex functions. Three of these proofs are reviewed in this section and the next. We have singled out those proofs that are of use in obtaining further information concerning optimal strategies (see Remark 1.4 below).

The first proof presented is based on the idea of separating convex sets by supporting hyperplanes. Specific geometric representations of optimal strategies are developed in the course of the argument. Similar representations will be used later in analyzing the nature of the sets of optimal strategies.

► **LEMMA 1.4.1.** If  $S^n$  denotes the collection of all vectors in  $E^n$  satisfying  $x_i \geq 0$  ( $i = 1, \dots, n$ ),  $\sum_{i=1}^n x_i = 1$ , and  $\mathbf{b} = \langle b_1, \dots, b_n \rangle$  is a fixed vector, then

$$\max_{x \in S^n} \sum_{i=1}^n x_i b_i = \max_i b_i. \quad (1.4.1)$$

*Proof.* Clearly,  $\max_i b_i = \sum_{j=1}^n (x_j \max_i b_i) \geq \sum_{j=1}^n x_j b_j$ . On the other hand,  $\max_i b_i = b_{i_0} \leq \max_{x \in S^n} \sum_{i=1}^n x_i b_i$ .

We next develop a geometrical interpretation for strategies and a schematic representation for the value of a game. To this end, we consider the pay-off matrix

$$\mathbf{A} = \begin{vmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{vmatrix} = \|\mathbf{a}_1, \dots, \mathbf{a}_m\|,$$

where  $\mathbf{a}_i$  stands for the column vector  $\langle a_{1i}, \dots, a_{ni} \rangle$ . The  $\mathbf{a}_i$ 's can be viewed as  $m$  points in  $n$ -dimensional space.

Let  $\Gamma$  denote the convex set in  $E^n$ -space spanned by the  $\mathbf{a}_i$ . Equivalently,  $\Gamma$  is the intersection of all convex sets containing all  $\mathbf{a}_i$ . Formally,

$$\Gamma = \left\{ \sum_{j=1}^m y_j \mathbf{a}_j \mid \sum_{j=1}^m y_j = 1, y_j \geq 0 \right\}.$$

*Admissible values.* We call a real number  $\lambda$  admissible if there exists a vector  $\mathbf{p} \in \Gamma$  such that  $\mathbf{p} \leq \boldsymbol{\lambda}$ . (Recall that  $\mathbf{p} \leq \boldsymbol{\lambda}$  means that every component  $p_i$  is smaller than or equal to  $\lambda$ ; see p. 10.)

If  $\Omega$  denotes the set of admissible numbers, then (i)  $\Omega$  is nonvoid since it contains  $N$ , provided  $N > a_{ij}$  for all  $i, j$ , and (ii)  $\Omega$  is bounded from below by  $-N$ , for  $N$  sufficiently large.

Let

$$\lambda_0 = \inf_{\lambda \in \Omega} \lambda.$$

In order to show that  $\lambda_0$  is admissible—i.e., that  $\lambda_0 \in \Omega$ —we select a sequence  $\{\lambda_n\}$  in  $\Omega$  such that  $\lambda_n \rightarrow \lambda_0$ .

To every  $\lambda_n$  there corresponds at least one vector  $\mathbf{q}_n$  in  $\Gamma$  such that  $\mathbf{q}_n \leq \boldsymbol{\lambda}_n$ . Since  $\Gamma$  is compact and the sequence  $\{\mathbf{q}_n\}$  must have at least one limit point  $\mathbf{q}_0$ , it follows readily that  $\mathbf{q}_0 \leq \boldsymbol{\lambda}_0$ , and thus  $\lambda_0$  is admissible.

Consider the translation of the negative orthant  $O_{\lambda_0}$  consisting of all vectors having no component greater than  $\lambda_0$ . The set  $O_{\lambda_0}$  touches  $\Gamma$  with no overlap. In fact,  $O_{\lambda_0} \cap \Gamma \supset \mathbf{q}_0$ , but if the interior of  $O_{\lambda_0}$  were to contain points of  $\Gamma$ , the meaning of  $\lambda_0$  would be contradicted.

*Supporting planes to  $O_{\lambda_0}$ .* We shall now characterize the form of all supporting planes to  $O_{\lambda_0}$  that are oriented so that their normals are directed into the half-space not containing  $O_{\lambda_0}$  (see also Lemma B.1.3 in Appendix B).

Let us consider a hyperplane defined by

$$\sum_{i=1}^n \mu_i \xi_i + \mu_0 = 0, \quad \sum_{i=1}^n |\mu_i| \neq 0 \quad (1.4.2)$$

whose normal is directed away from  $O_{\lambda_0}$ . This provides a supporting plane to  $O_{\lambda_0}$  if and only if for every  $\xi$  in  $O_{\lambda_0}$

$$\sum_{i=1}^n \mu_i \xi_i + \mu_0 \leq 0, \quad (1.4.3)$$

and there exists a  $\xi^0$  in  $O_{\lambda_0}$  for which equality holds. We now show that  $\mu_i$  can be characterized as follows:

$$\mu_i \geq 0 \quad (i = 1, \dots, n), \quad \sum_{i=1}^n \mu_i > 0, \quad \mu_0 = -\lambda_0 \sum_{i=1}^n \mu_i. \quad (1.4.4)$$

That  $\mu_i$  ( $1 \leq i \leq n$ ) is nonnegative is seen by applying (1.4.3) to the vector  $\xi = \langle \xi_1, \dots, \xi_n \rangle$  with components  $\xi_j = \lambda_0$  for  $j \neq i$ ,  $\xi_i = -N$ , and  $N$  sufficiently large. Since (1.4.2) defines a hyperplane, it follows that

$$\sum_{i=1}^n \mu_i > 0.$$

Finally, any supporting plane to  $O_{\lambda_0}$  clearly must contain the vertex point  $\xi = \langle \lambda_0, \lambda_0, \dots, \lambda_0 \rangle$ . Therefore,

$$O = \left( \sum_{i=1}^n \mu_i \right) \lambda_0 + \mu_0$$

as asserted. Conversely, any set of  $\mu_i$  satisfying the conditions of (1.4.4) defines a supporting hyperplane to  $O_{\lambda_0}$  whose normal is directed into the half-space not containing  $O_{\lambda_0}$ .

If we change notation from  $\mu_i$  to  $x_i$ , supporting planes to  $O_{\lambda_0}$  whose normal is directed away from  $O_{\lambda_0}$  can be characterized by the relation

$$\sum_{i=1}^n x_i \xi_i - \lambda_0 = 0,$$

where  $\sum_{i=1}^n x_i = 1$  and  $x_i \geq 0$ .

*Geometrical representation of strategies.* Strategies  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  in  $S^n$ , i.e.,

$$\sum_{i=1}^n x_i = 1, \quad x_i \geq 0,$$

generate supporting planes to  $O_{\lambda_0}$  that are defined by the equation

$$\sum_{i=1}^n x_i \xi_i - \lambda_0 = 0,$$

and conversely.

Strategies  $\mathbf{y} = \langle y_1, \dots, y_m \rangle$  in  $T^m$  for Player II correspond to points in  $\Gamma$ . To  $\mathbf{y}$  we associate the point  $\mathbf{p}$  with coordinates

$$p_i = \sum_{j=1}^m a_{ij} y_j.$$

Unfortunately, the correspondence of  $Y$  to  $\Gamma$  is not one to one, since subsets of  $Y$  map into points of  $\Gamma$  if and only if  $\mathbf{A}$  has rank less than  $m$ .

Finally, the expected pay-off  $A(\mathbf{x}, \mathbf{y})$  can be interpreted as the inner product of the normal to the plane determined by  $\mathbf{x}$ , and the point of  $\Gamma$  determined by  $\mathbf{y}$ . Explicitly,

$$(\mathbf{x}, \mathbf{Ay}) = \sum_{i=1}^n \sum_{j=1}^m x_i a_{ij} y_j.$$

► **THEOREM 1.4.1 (Min-Max Theorem).** If  $\mathbf{x}$  and  $\mathbf{y}$  range over  $S^n$  and  $T^m$ , respectively, then

$$\min_{\mathbf{y}} \max_{\mathbf{x}} A(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x}} \min_{\mathbf{y}} A(\mathbf{x}, \mathbf{y}) = v.$$

*Proof.* We construct  $O_{\lambda_0}$  and  $\Gamma$  as described above.

Since  $O_{\lambda_0}$  and  $\Gamma$  are nonoverlapping convex sets by Lemma B.1.2 in Appendix B, there exists a separating plane. Since  $O_{\lambda_0}$  and  $\Gamma$  touch, the separating plane must also be a supporting plane, which may be oriented so that its normal is directed away from  $O_{\lambda_0}$ .

Let the direction numbers of its normal be  $x_i^0$ . Then, since

$$\sum_{i=1}^n x_i^0 = 1$$

and  $x_i^0 \geq 0$ , as required by (1.4.4),

$$\sum_{i=1}^n x_i^0 \xi_i - \lambda_0 \leq 0 \quad \text{for all } \xi \in O_{\lambda_0},$$

and

$$\sum_{i=1}^n x_i^0 \xi_i - \lambda_0 \geq 0 \quad \text{for all } \xi \in \Gamma.$$

If  $\mathbf{y}$  is any given strategy, then the vector  $\xi$ , where

$$\xi_i = \sum_{j=1}^m a_{ij} y_j,$$

is in  $\Gamma$ , and thus

$$A(\mathbf{x}^0, \mathbf{y}) = \sum_{i=1}^n x_i^0 \sum_{j=1}^m a_{ij} y_j \geq \lambda_0 \quad (1.4.5)$$

for any strategy  $\mathbf{y}$ .

If  $\mathbf{y}^0$  denotes a strategy corresponding to a point  $\Gamma$  which also belongs to  $O_{\lambda_0}$ , then

$$\sum_{j=1}^m a_{ij} y_j^0 \leq \lambda_0 \quad \text{for all } i, \quad (1.4.6)$$

and hence  $A(\mathbf{x}, \mathbf{y}^0) \leq \lambda_0$  for all  $\mathbf{x}$  strategies (by Lemma 1.4.1).

The proof of the theorem is complete by virtue of Theorem 1.3.1.

*Remark 1.4.* We have observed that strategies for I can be viewed as supporting planes, and strategies for II as points in  $\Gamma$ . We further see that optimal strategies for I correspond to separating planes between  $O_{\lambda_0}$  and  $\Gamma$ , and that optimal strategies for II correspond to those points of  $\Gamma$  which have contact with  $O_{\lambda_0}$ . This geometric characterization of optimal strategies will be exploited in Chapter 3.

The interpretations used in the arguments of Theorem 1.4.1 are unsymmetrical, in that the  $\mathbf{x}$  strategies correspond to hyperplanes and the  $\mathbf{y}$

strategies to points of Euclidean  $n$ -space. However, a completely parallel analysis can be developed which associates  $\mathbf{y}$  strategies with hyperplanes and  $\mathbf{x}$  strategies with points. In this analysis the analog of  $\Gamma$  is the convex set spanned by the row vectors of  $\mathbf{A}$ , and so on.

It is often of interest to know when there exist optimal pure strategies. This will be the case when there exist pure strategies  $\{\alpha_{i_0}, \beta_{j_0}\}$  such that

$$\begin{aligned} a_{i_0, j} &\geq a_{i_0, j_0} && \text{for all } j, \\ a_{i_0, j_0} &\geq a_{i, j_0} && \text{for all } i. \end{aligned} \tag{1.4.7}$$

The first inequality implies that for any mixed strategy  $\mathbf{y}$ ,

$$\sum_{j=1}^m a_{i_0, j} y_j \geq a_{i_0, j_0} = v,$$

or  $A(\alpha_{i_0}, \mathbf{y}) \geq v$ . Similarly, we obtain  $v \geq A(\mathbf{x}, \beta_{j_0})$ , and thus  $\alpha_{i_0}$  and  $\beta_{j_0}$  are optimal strategies for Players I and II, respectively.

Where pure optimal strategies exist, the matrix  $\mathbf{A}$  is said to have a *saddle point* at  $(i_0, j_0)$ . A saddle point is an element of a matrix which is both a minimum of its row and a maximum of its column.

**\*1.5 General min-max theorem.** In this section two alternative proofs of the min-max theorem for general finite games are given. The first proof is based on the Kakutani fixed-point theorem. The second proof relies only on the properties of convex functions.

► **THEOREM 1.5.1.** Let  $f(\mathbf{x}, \mathbf{y})$  be a real-valued function of two variables  $\mathbf{x}$  and  $\mathbf{y}$  which traverse  $C$  and  $D$ , respectively, where both  $C$  and  $D$  are closed, bounded, convex sets. If  $f$  is continuous, convex in  $\mathbf{y}$  for each  $\mathbf{x}$ , and concave in  $\mathbf{x}$  for each  $\mathbf{y}$ , then  $\min_{\mathbf{y}} \max_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x}} \min_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ .

*Proof.* For each  $\mathbf{x}$ , define  $\phi_x(\mathbf{y}) = f(\mathbf{x}, \mathbf{y})$ . Let

$$B_x = \{\mathbf{y} | \phi_x(\mathbf{y}) = \min_{\mathbf{y} \in D} \phi_x(\mathbf{y})\}.$$

It is easily verified that  $B_x$  is nonvoid, closed, and convex. The fact that  $B_x$  is convex is a consequence of the convexity of the function  $f$  in the variable  $\mathbf{y}$ . For each  $\mathbf{y}$ , define  $\psi_y(\mathbf{x}) = f(\mathbf{x}, \mathbf{y})$ . Let

$$A_y = \{\mathbf{x} | \psi_y(\mathbf{x}) = \max_{\mathbf{x} \in C} \psi_y(\mathbf{x})\}.$$

Again,  $A_y$  is nonvoid, closed, and convex.

---

\* Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

Let the set  $E$  be the direct product,  $E = C \times D$ , of the convex sets  $C$  and  $D$ , and recall that the direct product of two convex sets is itself convex.

Let  $g$  be the mapping which maps the point  $\{\mathbf{x}, \mathbf{y}\} \in E$  into the set  $(A_y \times B_x)$ . Since  $A_y$  and  $B_x$  are nonvoid, closed, convex subsets of  $C$  and  $D$ , respectively,  $(A_y \times B_x)$  is a nonvoid, closed, convex subset of  $E$ . Hence, condition (i) of the Kakutani theorem is fulfilled (see Section C.2 in the Appendix). It is also easy to verify that the continuity requirement of the Kakutani theorem is satisfied. Therefore, invoking the Kakutani theorem, there exists an  $\{\mathbf{x}^0, \mathbf{y}^0\} \in (A_{y^0} \times B_{x^0})$ , that is,

$$\begin{aligned} f(\mathbf{x}^0, \mathbf{y}^0) &\geq f(\mathbf{x}, \mathbf{y}^0) \quad \text{for all } \mathbf{x} \in C, \\ f(\mathbf{x}^0, \mathbf{y}^0) &\leq f(\mathbf{x}^0, \mathbf{y}) \quad \text{for all } \mathbf{y} \in D. \end{aligned}$$

It then follows from Corollary 1.3.1 that

$$\min_{y \in D} \max_{x \in C} f(\mathbf{x}, \mathbf{y}) = \max_{x \in C} \min_{y \in D} f(\mathbf{x}, \mathbf{y}). \quad (1.5.1)$$

Whenever (1.5.1) is satisfied, we say that the game defined by  $(f, C, D)$  possesses a value or, alternatively, that the game is determined.

The other proof proceeds as follows. Assume first that  $f(\mathbf{x}, \mathbf{y})$  is a strictly convex function of  $\mathbf{y}$ , for fixed  $\mathbf{x}$ , and  $f(\mathbf{x}, \mathbf{y})$  is a strictly concave function of  $\mathbf{x}$ , for fixed  $\mathbf{y}$ . By virtue of the strict convexity, there exists a unique  $\mathbf{y}(\mathbf{x})$  for which

$$f[\mathbf{x}, \mathbf{y}(\mathbf{x})] = \min_y f(\mathbf{x}, \mathbf{y}) = m(\mathbf{x}).$$

From the uniform continuity of  $f$  and the uniqueness of  $\mathbf{y}(\mathbf{x})$  it follows readily that  $m(\mathbf{x})$  and  $\mathbf{y}(\mathbf{x})$  are continuous. Also, since the minimum of a family of concave functions remains concave,  $m(\mathbf{x})$  is concave. Let  $\mathbf{x}^*$  be a point where

$$m(\mathbf{x}^*) = \max_x m(\mathbf{x}) = \max_x \min_y f(\mathbf{x}, \mathbf{y}).$$

For any  $\mathbf{x}$  in  $C$  and any  $t$  in  $0 < t < 1$ , we have by the hypothesis that

$$\begin{aligned} f[(1 - t)\mathbf{x}^* + t\mathbf{x}, \mathbf{y}] &> (1 - t)f(\mathbf{x}^*, \mathbf{y}) + tf(\mathbf{x}, \mathbf{y}) \\ &\geq (1 - t)m(\mathbf{x}^*) + tf(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Put  $\mathbf{y} = \mathbf{y}[(1 - t)\mathbf{x}^* + t\mathbf{x}] = \tilde{\mathbf{y}}$ , so that

$$m[(1 - t)\mathbf{x}^* + t\mathbf{x}] \geq (1 - t)m(\mathbf{x}^*) + tf(\mathbf{x}, \tilde{\mathbf{y}}).$$

But since  $m(\mathbf{x}^*) \geq m(\mathbf{x})$  for all  $\mathbf{x}$  in  $C$ , it follows that

$$m(\mathbf{x}^*) \geq m[(1 - t)\mathbf{x}^* + t\mathbf{x}],$$

and hence

$$f(\mathbf{x}, \mathbf{y}) \leq m(\mathbf{x}^*) = f[\mathbf{x}^*, \mathbf{y}(\mathbf{x}^*)].$$

Now let  $t \rightarrow 0$ , so that  $(1 - t)\mathbf{x}^* + t\mathbf{x} \rightarrow \mathbf{x}^*$  and  $\mathbf{y} \rightarrow \mathbf{y}(\mathbf{x}^*)$ ; from this we deduce that

$$f[\mathbf{x}, \mathbf{y}(\mathbf{x}^*)] \leq f[\mathbf{x}^*, \mathbf{y}(\mathbf{x}^*)] \quad (1.5.2)$$

for any  $\mathbf{x}$ . If we let  $\mathbf{y}(\mathbf{x}^*) = \mathbf{y}^*$  and  $v = f(\mathbf{x}^*, \mathbf{y}^*)$ , then it is clear from (1.5.2) and the construction of  $\mathbf{y}(\mathbf{x})$  that

$$f(\mathbf{x}, \mathbf{y}^*) \leq v \leq f(\mathbf{x}^*, \mathbf{y}).$$

This is equivalent to the min-max theorem. It remains only to relax the conditions of strictness imposed above. To this end, define

$$f_\epsilon(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) - \epsilon f(\mathbf{x}) + \epsilon g(\mathbf{y}),$$

where

$$f(\mathbf{x}) = \sum_{i=1}^n x_i^2, \quad g(\mathbf{y}) = \sum_{j=1}^m y_j^2.$$

Each  $f_\epsilon$  satisfies the strict convexity-concavity requirements used in the previous argument, and consequently

$$f(\mathbf{x}, \mathbf{y}_\epsilon) - \epsilon f(\mathbf{x}) \leq f_\epsilon(\mathbf{x}, \mathbf{y}_\epsilon) \leq v_\epsilon \leq f_\epsilon(\mathbf{x}_\epsilon, \mathbf{y}) \leq f(\mathbf{x}_\epsilon, \mathbf{y}) + \epsilon g(\mathbf{y})$$

for all  $\{\mathbf{x}, \mathbf{y}\}$  in  $C \times D$ . Take a sequence of  $\epsilon$ 's approaching zero and consider a subsequence  $\mathbf{x}_\epsilon \rightarrow \mathbf{x}^*$ ,  $\mathbf{y}_\epsilon \rightarrow \mathbf{y}^*$ ,  $v_\epsilon \rightarrow v^*$ . We obtain in the limit

$$f(\mathbf{x}, \mathbf{y}^*) \leq v^* \leq f(\mathbf{x}^*, \mathbf{y}),$$

and the proof is complete.

### 1.6 Problems.

1. "Stone, Paper, and Scissors." Both players must simultaneously name one of these objects; "paper" defeats "stone," "stone" defeats "scissors," and "scissors" defeats "paper." The player who chooses the winning object wins one unit; if both players choose the same object, the game is a draw. Set up the matrix for this game and describe the optimal strategies.

2. Two cards marked Hi and Lo are placed in a hat. Player I draws a card at random and inspects it. He may pass immediately, in which case he pays Player II the amount  $a > 0$ ; or he may bet, in which case Player II may either pass, paying amount  $a$  to Player I, or call. If he calls, he

receives from or pays to Player I the amount  $b > a$ , according as Player I holds Lo or Hi. Set up the game and solve by the methods of Theorem 1.4.1.

3. Player I draws a card Hi or Lo from a hat. He then has the option of either betting or passing. If he bets, Player II may either fold, in which case he pays Player I an amount  $a$ , or call, in which case he wins or loses  $b > a$  according as Player I holds Hi or Lo. If Player I passes initially, Player II draws Hi or Lo from a new hat. Player II then has the symmetrical choice of actions, to which Player I must respond with the corresponding pay-off. If Player II passes on the second round, the pay-off is zero. Describe the game and the strategy spaces.

4. Each player must choose a number between 1 and 9. If a player's choice is one unit higher than his opponent's, he loses two dollars; if it is at least two units higher, he wins one dollar; if both players choose the same number, the game is a draw. Set up the matrix for this game.

5. Show that the  $n \times n$  matrix  $\|a_{ij}\|$  with  $a_{ij} = i - j$  has a saddle point, and describe the corresponding optimal strategies.

6. Use Theorem 1.4.1 to solve the game with matrix

$$\begin{vmatrix} 3 & -1 & 3 & 7 \\ -1 & 9 & 3 & 0 \end{vmatrix}.$$

7. Two games have matrices  $\mathbf{A} = \|a_{ij}\|$  and  $\mathbf{B} = \|b_{ij}\|$ , respectively; the value of the first game is  $v$ , and  $b_{ij} = a_{ij} + a$ . Show that the value of  $\mathbf{B}$  is  $v + a$ , and that the optimal strategies are the same for both games.

8. Show that each of the two matrices

$$\mathbf{A}_1 = \begin{vmatrix} 0 & x \\ 1 & 2 \end{vmatrix}, \quad \mathbf{A}_2 = \begin{vmatrix} 2 & 1 \\ x & 0 \end{vmatrix}$$

has a saddle point, and find values  $x_1$  and  $x_2$  such that

$$v(\mathbf{A}_1 + \mathbf{A}_2) < v(\mathbf{A}_1) + v(\mathbf{A}_2)$$

and

$$v(\mathbf{A}_1 + \mathbf{A}_2) > v(\mathbf{A}_1) + v(\mathbf{A}_2),$$

respectively.

9. Give an example to show that

$$(\mathbf{x}^*, \mathbf{Ay}^*) = \min_{y \in Y} \max_{x \in X} (\mathbf{x}, \mathbf{Ay}) = \max_{x \in X} \min_{y \in Y} (\mathbf{x}, \mathbf{Ay}),$$

is not sufficient proof that  $\mathbf{x}^*$  and  $\mathbf{y}^*$  are optimal.

10. Let  $\mathbf{A} = \|a_{ij}\|$  be an  $n \times m$  matrix. Show that either there exists

$$\mathbf{u} = \langle u_1, \dots, u_n \rangle \quad \left( \text{with } u_i \geq 0, \sum_{i=1}^n u_i = 1 \right)$$

such that

$$\sum_{i=1}^n u_i a_{ij} \geq 0 \quad \text{for all } j,$$

or there exists

$$\mathbf{w} = \langle w_1, \dots, w_m \rangle \quad \left( \text{with } w_j \geq 0, \sum_{j=1}^m w_j = 1 \right)$$

such that

$$\sum_{j=1}^m a_{ij} w_j < 0 \quad \text{for all } i.$$

11. Let  $\mathbf{B} = \|b_{ij}\|$  be a fixed  $n \times m$  game, and let  $f(t)$  be a convex increasing function of  $t$ . Let  $\Gamma$  be the game with strategies

$$\mathbf{x} = \langle x_1, \dots, x_n \rangle, \quad \mathbf{y} = \langle y_1, \dots, y_m \rangle$$

and pay-off

$$\sum_{i=1}^n x_i f \left( \sum_{j=1}^m b_{ij} y_j \right).$$

Show that the optimal strategies for  $\mathbf{B}$  are also optimal for the new game.

12. Show that a game  $\sum x_i a_{ij} y_j$  played over closed sets of strategies  $S$  and  $T$  has a value if and only if the game played over the convex hulls of  $S$  and  $T$  has two optimal strategies  $\mathbf{x}^* \in S$  and  $\mathbf{y}^* \in T$ .

13. Let  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{D}$  be matrix games of orders  $n_1 \times m_1$ ,  $n_1 \times m_2$ ,  $n_2 \times m_2$  and values  $a$ ,  $b$ ,  $d$ , respectively. Let  $\{\mathbf{x}^0, \mathbf{y}^0\}$  and  $\{\mathbf{x}^*, \mathbf{y}^*\}$  be optimal for  $\mathbf{A}$  and  $\mathbf{D}$ , respectively, and let  $\langle \xi, 1 - \xi \rangle$  and  $\langle \eta, 1 - \eta \rangle$  be optimal for the game

$$\begin{array}{c|cc} & a & b \\ \hline 0 & & d \end{array}.$$

Show that if  $a < 0 < d$ , the strategies

$$\mathbf{x} = \langle \xi x_1^0, \dots, \xi x_{n_1}^0, (1 - \xi)x_{n_1+1}^*, \dots, (1 - \xi)x_{n_1+m_2}^* \rangle,$$

$$\mathbf{y} = \langle \eta y_1^0, \dots, \eta y_{m_1}^0, (1 - \eta)y_{m_1+1}^*, \dots, (1 - \eta)y_{m_1+m_2}^* \rangle$$

are optimal for the game

$$\begin{array}{c|cc} \mathbf{A} & \mathbf{B} \\ \hline 0 & \mathbf{D} \end{array}.$$

14. Show that the value of a matrix game is a nondecreasing continuous function of the components of the matrix.

15. Let  $X$  and  $Y$  be convex compact sets, and let  $f(\mathbf{x}, \mathbf{y})$  be concave in  $\mathbf{x}$  and convex in  $\mathbf{y}$ . Show that for any finite set  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_m$  there exists a pair  $\{\mathbf{x}_0, \mathbf{y}_0\}$  such that

$$\max_{1 \leq i \leq n} f(\mathbf{x}_i, \mathbf{y}_0) \leq \min_{1 \leq j \leq m} f(\mathbf{x}_0, \mathbf{y}_j).$$

16. Let  $X$  and  $Y$  be convex sets, with  $Y$  also compact. Let  $f(\mathbf{x}, \mathbf{y})$  be a function defined on  $X \times Y$  with the property that  $f$  is lower-semicontinuous in  $\mathbf{y}$ , concave in  $\mathbf{x}$ , and convex in  $\mathbf{y}$ . Show that for each real  $\alpha$ , either there exists  $\mathbf{y}_0$  in  $Y$  such that  $f(\mathbf{x}, \mathbf{y}_0) \leq \alpha$  for all  $\mathbf{x}$  in  $X$ , or there exists  $\mathbf{x}_0$  in  $X$  such that  $f(\mathbf{x}_0, \mathbf{y}) > \alpha$  for all  $\mathbf{y}$  in  $Y$ .

17. Show that if every  $2 \times 2$  submatrix of a matrix  $\mathbf{A}$  has a saddle point, then  $\mathbf{A}$  has a saddle point.

18. Let  $\mathbf{A}$  be a nonsingular  $n \times m$  matrix with  $m \geq 3$ . Show that if every  $n \times m - 1$  submatrix has a strict saddle point, then  $\mathbf{A}$  has a saddle point.

19. Let  $X$  and  $Y$  denote polyhedral convex sets in  $E^n$  and  $E^m$  space, respectively. Let  $\mathbf{A}$  be an  $n \times m$  matrix. Define

$$v_2 = \inf_{\mathbf{y} \in Y} \sup_{\mathbf{x} \in X} (\mathbf{x}, \mathbf{A}\mathbf{y}), \quad v_1 = \sup_{\mathbf{x} \in X} \inf_{\mathbf{y} \in Y} (\mathbf{x}, \mathbf{A}\mathbf{y}).$$

Prove that if  $v_1 \neq v_2$ , then  $v_1 = -\infty$  and  $v_2 = +\infty$ . Construct examples satisfying  $v_1 = -\infty$ ;  $v_2 = +\infty$ ;  $v_1 = v_2 = +\infty$ ;  $v_1 = v_2 = -\infty$ . (A polyhedral convex set is defined as the intersection of a finite number of closed half spaces.)

## NOTES AND REFERENCES TO CHAPTER 1

**§ 1.1.** Several popular and entertaining discussions of discrete game theory and its applications are available for the nonmathematically-oriented reader, notably Williams [254] and McDonald [176]. A comprehensive critical survey of the concepts of game theory, utility theory, and related subjects is contained in Luce and Raiffa [169].

Historically, the modern mathematical approach to game theory is attributed to von Neumann in his papers of 1928 [238] and 1937 [239], although some game problems were dealt with earlier by Borel [39]. The primary stimulation for the present vigorous research in game theory was the publication in 1944 of the fundamental book on games and economic behavior by von Neumann and Morgenstern [242]. For introductory mathematical expositions in game theory, the reader is directed to McKinsey [179], Gale [96], and Vajda [236].

To the more advanced student we recommend the volumes edited by Kuhn and Tucker [161, 162]. Another volume of studies in game theory has been compiled by Dresher, Tucker, and Wolfe [72]. To the student interested primarily in  $n$ -person game theory we call attention to Shapley [216] and a volume edited by Luce and Tucker [170].

Several texts on related subjects incorporate some aspects of game theory. These include Blackwell and Girshick [30], Churchman, Ackoff, and Arnoff [50], McCloskey and Trefethen [175], and Allen [2].

Luce and Raiffa, in addition to their discussions of the underlying concepts of games and decisions, provide in eight appendixes a moderate survey of several of the techniques of the subject.

A handbook of important results in game theory has been compiled by Dresher [66].

Although a knowledge of utility theory is not essential to the analytical study of the structure of games and their solutions, a few words on the subject may be helpful.

Von Neumann and Morgenstern in their classic book introduced an axiom system satisfied by a preference pattern  $\succ$  defined on a finite set of outcomes and the space  $\mathcal{P}$  of probability distributions of outcomes. That is to say, of any two probability distributions  $P_1$  and  $P_2$  on the space of outcomes, the preference relation selects one as preferable to the other. (In symbols, either  $P_1 \succ P_2$  or  $P_2 \succ P_1$ .)

The axioms, stated loosely, are as follows:

I. If  $P_1 \succ P_2$  and  $Q_1 \succ Q_2$ , then any probability mixture of  $P_1$  and  $Q_1$  is preferred to the same probability mixture of  $P_2$  and  $Q_2$ .

II. (Continuity Axiom.) There is no probability distribution on the space of outcomes that is infinitely desirable or infinitely undesirable. Formally, we may express this axiom as follows: If  $P_1 \prec P_2 \prec P_3$ , there are numbers  $\lambda < 1$  and  $\mu > 0$  such that

$$\lambda P_1 + (1 - \lambda) P_3 \prec P_2, \quad \mu P_1 + (1 - \mu) P_3 \succ P_2.$$

When the preference relation obeys these axioms, it can be expressed as a numerical utility function  $U$  defined for probability distributions on the set of outcomes. Furthermore,  $U$  is linearly convex with respect to the operation of forming convex combinations of elements of  $\mathcal{P}$ . It is also shown by these authors that  $U$  is uniquely determined up to a positive linear affine transformation. That is, if  $U$  is a utility function, then  $aU + b$  (with  $a > 0$  and  $b$  real) is the only other possible utility function. The proof is given in Chapter i of [242].

Blackwell and Girshick [30] have extended this result to the case in which the set of outcomes may contain a countable number of elements. Luce and Raiffa [169] subject the von Neumann and Morgenstern axioms to a penetrating review. We recommend their critical discussion to the reader interested in the foundations of game theory.

In the classical approach to the theory of games, it is necessary when defining a strategy to specify decisions for both players which appropriately take account of the complete present state of information. Because of chance moves which

may occur in the course of the play the outcomes of a pair of strategies may take the form of a random variable with a known probability distribution. Generally, in choosing among strategies, it is necessary to specify a preference relation which chooses among probability distributions over the set of outcomes. If the von Neumann and Morgenstern axioms are valid for this preference relation, then the preference associated with any pair of pure strategies can be expressed by a number obtained by applying utility notions to distributions over outcomes. In this way the pay-off function can be so constructed as to be consistent with the preference pattern. Moreover, the pay-off operates additively (i.e., is linearly convex) when computing expected values with respect to probability distributions on the set of pure strategies. Hence the formula (1.1.1) obtains.

**§ 1.2.** The poker example of this section was first introduced and studied by Kuhn [156]. A general technique for the analysis of parlor games is developed in Volume II Chapter 9 of the present work.

**§ 1.3.** The concept of the upper and lower values was proposed by von Neumann, who refers to corresponding situations as the majorant game and minorant game, respectively.

**§ 1.4.** Proofs of the fundamental min-max theorems have been given by von Neumann [239], Ville [237], Loomis [168], Karlin [133], Fan [77], Glicksberg [108], and others. All proofs except Loomis' utilize in essence some form of convexity and compactness of the strategy spaces.

**§ 1.5.** The first proof in this section is due to Kakutani [130], who obtained, in the course of establishing the min-max theorem, a generalization of the classical Brouwer fixed-point theorem. Kakutani's method depends on certain convexity properties of the mapping function considered. Eilenberg and Montgomery [73] and Begle [18] have shown that this generalized fixed-point theorem is valid when the requirement of simple connectedness for certain sets is substituted for convexity. Further extensions along these lines have been given by Debreu [61]. The second proof in this section was proposed by Shiffman [221].

**§ 1.6.** Problems 2 and 3 are based on Borel [39]. The statement of Problem 12 was first observed by I. Glicksberg. The conclusions of Problems 15 and 16 were used by Fan [77]. Problems 17 and 18 were suggested by L. Shapley. Problem 19 is due to Wolfe [255].

## CHAPTER 2

### THE NATURE OF OPTIMAL STRATEGIES FOR MATRIX GAMES

In this chapter we develop some of the fundamental working tools needed in the analysis of matrix games. Particular attention is directed to characterizing certain basic extreme-point optimal strategies. Since the extreme-point optimal strategies span the set of all optimal strategies, knowing the former is essentially equivalent to knowing the latter; i.e., every optimal strategy is a convex combination of extreme-point optimal strategies. The characterization of the extreme-point optimal strategies can be used to set up a computational scheme for obtaining the full set of optimal strategies.

The concept of dominance among strategies, which is introduced in Section 2.2, enables us to eliminate obviously poor strategies from consideration in the early stages of the consideration of the game. The dominance principle is especially useful in solving games explicitly, since many seemingly large-scale games can be reduced to small matrix games by excluding all dominated strategies. In this chapter four types of dominance are delimited, and the techniques associated with them are developed.

An important subclass of games called completely mixed games is also introduced in this chapter. These are games in which all pure strategies have relevance—i.e., have positive probability—in every optimal strategy. Several examples of such games are considered.

At the close of the chapter, symmetric games are discussed. A symmetric game is the equivalent of a “fair game,” one in which the opportunities and possible pay-offs are equal for both players.

**2.1 Properties of optimal strategies.** In this section we establish two basic properties of optimal strategies for matrix games. These properties will be used in subsequent sections.

► **LEMMA 2.1.1.** The set of optimal strategies for each player is a convex closed set.

*Proof.* We shall give the proof for Player I only. If  $v$  is the value of the game and  $\mathbf{x}_0$  and  $\mathbf{x}'_0$  are optimal strategies for Player I, then  $A(\mathbf{x}_0, \mathbf{y}) \geq v$  and  $A(\mathbf{x}'_0, \mathbf{y}) \geq v$  for all  $\mathbf{y}$ . Thus

$$A[\lambda \mathbf{x}_0 + (1 - \lambda) \mathbf{x}'_0, \mathbf{y}] = \lambda A(\mathbf{x}_0, \mathbf{y}) + (1 - \lambda) A(\mathbf{x}'_0, \mathbf{y}) \geq v$$

for all  $\mathbf{y}$ , and therefore  $\lambda \mathbf{x}_0 + (1 - \lambda) \mathbf{x}'_0$  is contained in the set  $X^0$  of optimal strategies.

That  $X^0$  is closed follows from the continuity of  $A(\mathbf{x}, \mathbf{y})$  as a function of  $\mathbf{x}$ .

►LEMMA 2.1.2. If  $\mathbf{y}^0$  is an optimal mixed strategy for Player II and  $y_{j_0}^0 > 0$ , then every optimal  $\mathbf{x}$  strategy for Player I must have the property that

$$A(\mathbf{x}, j_0) = \sum_{i=1}^n x_i a_{ij_0} = v.$$

The corresponding result is true for the other player.

*Proof.* Let  $\mathbf{x}$  be an optimal strategy.

Since  $\langle 0, \dots, 1, 0, \dots, 0 \rangle$ , where 1 occurs only in the  $r$ th component, is a possible  $\mathbf{y}$  strategy, it follows that

$$\sum_{i=1}^n x_i a_{ir} \geq v \quad \text{for all } r. \quad (2.1.1)$$

Suppose that strict inequality holds for  $r = j_0$ . Then

$$y_{j_0}^0 \sum_{i=1}^n x_i a_{ij_0} > v y_{j_0}^0,$$

since  $y_{j_0}^0 > 0$ , while for  $r \neq j_0$ ,

$$y_r^0 \sum_{i=1}^n x_i a_{ir} \geq v y_r^0.$$

Hence

$$\sum_{r=1}^m y_r^0 \sum_{i=1}^n x_i a_{ir} > v \sum_{r=1}^m y_r^0 = v.$$

This contradicts the assumption that  $\mathbf{y}^0$  is optimal. Consequently,

$$\sum_{i=1}^n x_i a_{ij_0} = v \quad \text{for all optimal } \mathbf{x}.$$

For later reference we introduce here some terminology. If  $x_i$ , the  $i$ th component of a strategy  $\mathbf{x}$ , is positive, the corresponding pure strategy  $i$  is said to be relevant to, essential to, or employed with positive probability in the strategy  $\mathbf{x}$ . In referring to a pure strategy, we use the terms “ $i$ th row” and “ $i$ th pure strategy” interchangeably.

A row  $i$  such that

$$\sum_{j=1}^n a_{ij} y_j = v$$

is said to be an equalizer with respect to the strategy  $\mathbf{y}$ . In this context Lemma 2.1.2 may be restated as follows: If  $i$  is relevant to an optimal strategy  $\mathbf{x}^0$ , then  $i$  is an equalizer with respect to every optimal  $\mathbf{y}$  strategy.

The corresponding terminology will be employed with reference to  $\mathbf{y}$  strategies and columns. An optimal strategy will also be referred to as a solution to the game.

Finally, we shall frequently use the notation  $\mathbf{A}\mathbf{y} = \mathbf{v}$ , where  $\mathbf{A}$  is a matrix,  $\mathbf{y}$  is a vector, and  $\mathbf{v}$  is an ordinary scalar which in this context is to be interpreted as the vector  $\langle v, v, v, \dots, v \rangle$  with the appropriate number of components. We shall also use the analogous expressions  $\mathbf{A}\mathbf{y} \leq \mathbf{v}$  and  $\mathbf{x}\mathbf{A} \geq \mathbf{v}$ .

**2.2 Types of strict dominance.** The matrix associated with a game can often be reduced in size by the use of the dominance principle. When this principle is used, the strategies which are optimal with respect to the smaller matrix are also optimal with respect to the larger. Thus the concept of dominance can be used to simplify the problem of obtaining optimal strategies. Specific applications of this principle are given in Chapter 4.

The dominance principle states that if one strategy open to a player is better than another regardless of which strategy his opponent uses, then the inferior strategy may be disregarded in the search for optimal strategies. This principle and its applications are set forth explicitly in the following four theorems. These theorems are stated and proved for row dominance only; the analogous results for column dominance are stated without proof.

► **DEFINITION 2.2.1.** An  $n$ -tuple  $\alpha = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$  strictly dominates a second  $n$ -tuple  $\beta = \langle \beta_1, \dots, \beta_n \rangle$  if  $\alpha_i > \beta_i$  for  $i = 1, 2, \dots, n$ .

► **THEOREM 2.2.1.** If the  $j$ th row of the pay-off matrix  $\mathbf{A}$  strictly dominates the  $i$ th row, the  $i$ th row can be deleted from the matrix without changing the set of optimal strategies for Player I.

*Proof.* For optimal  $\mathbf{y}^0$

$$v \geq \sum_k a_{jk} y_k^0 > \sum_k a_{ik} y_k^0,$$

and by Lemma 2.1.2 no optimal  $\mathbf{x}^0$  can have  $x_i^0 > 0$ .

► **THEOREM 2.2.1a.** If the  $j$ th column strictly dominates the  $i$ th column, the  $j$ th column can be deleted without changing the set of optimal strategies for Player II.

► **THEOREM 2.2.2.** If the  $i$ th row is strictly dominated by a convex combination of the other rows, the  $i$ th row can be deleted.

*Proof.* The hypothesis of the theorem means that there exists a strategy  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  with  $x_i = 0$  such that

$$a_{ij} < \sum_{k=1}^n x_k a_{kj} \quad \text{for } j = 1, \dots, m. \quad (2.2.1)$$

If  $\mathbf{y}^0$  is optimal for Player II, it follows that

$$\sum_{j=1}^m a_{ij} y_j^0 < (\mathbf{x}, \mathbf{A}\mathbf{y}^0) \leq v.$$

The last inequality, in conjunction with Lemma 2.1.2, implies that the  $i$ th row cannot be relevant to any optimal  $\mathbf{x}$  strategy, and hence the conclusion of the theorem is established.

The condition “without changing the set of optimal strategies” is of course implied in this and the following theorems.

► **THEOREM 2.2.2a.** If the  $i$ th column strictly dominates some convex combination of the other columns, the  $i$ th column can be deleted.

► **THEOREM 2.2.3.** If a convex combination of rows in a set  $R_1$  strictly dominates a convex combination of rows in a set  $R_2$ , there exists a row in  $R_2$  that can be deleted.

*Proof.* Let  $\mathbf{x}^1$  be the convex combination of rows in  $R_1$  that strictly dominates the convex combination  $\mathbf{x}^2$  of rows in  $R_2$ ; that is,  $\mathbf{x}^1 \mathbf{A} \gg \mathbf{x}^2 \mathbf{A}$ .

Let us assume to the contrary that every row in  $R_2$  is employed with positive probability in some optimal strategy. Then we know from Lemma 2.1.2 that for  $\mathbf{y}^0$  optimal, each of the components in  $R_2$  of  $\mathbf{A}\mathbf{y}^0$  equals  $v$ . Thus we have  $v \geq (\mathbf{x}^1, \mathbf{A}\mathbf{y}^0) > (\mathbf{x}^2, \mathbf{A}\mathbf{y}^0) = v$ , which is impossible. Hence some row of  $R_2$  may be deleted; i.e., there is a row such that the component of any optimal  $\mathbf{x}$  for that row is 0.

► **THEOREM 2.2.3a.** If a convex combination of columns in a set  $C_1$  strictly dominates a convex combination of columns in a set  $C_2$ , there exists a column in  $C_1$  that may be deleted.

► **THEOREM 2.2.4.** Let

$$\mathbf{A} = \begin{vmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{vmatrix}.$$

If every column of  $\mathbf{A}_2$  strictly dominates some convex combination of the columns of  $\mathbf{A}_1$ , and if every row of  $\mathbf{A}_3$  is strictly dominated by some convex combination of the rows of  $\mathbf{A}_1$ , the submatrices  $\mathbf{A}_2$ ,  $\mathbf{A}_3$ , and  $\mathbf{A}_4$  may be deleted.

*Proof.* Suppose that row  $i$  of  $\mathbf{A}_3$  is employed with probability  $\lambda$  in some optimal strategy  $\mathbf{x}^*$ . Since some convex combination of the rows of  $\mathbf{A}_1$  strictly dominates  $i$ , we can spread the probability  $\lambda$  among these rows of  $\mathbf{A}_1$  proportionally to the weightings used in forming the convex combination and obtain a new strategy  $\mathbf{x}^0$ . We perform this construction for every row  $i$  of  $\mathbf{A}_3$  which is relevant to the optimal strategy  $\mathbf{x}^*$ . The final strategy obtained is denoted by  $\mathbf{x}'$ . Clearly, the expected yield when Player I uses  $\mathbf{x}'$  and Player II uses a pure strategy corresponding to a column of  $\mathbf{A}_1$  is strictly greater than the yield for the same column when Player I uses  $\mathbf{x}^*$ . In symbols,

$$(\mathbf{x}', \mathbf{A}\beta_j) > (\mathbf{x}^*, \mathbf{A}\beta_j) \geq v \quad (2.2.2)$$

for any pure strategy  $\beta_j$  with  $j$  in  $\mathbf{A}_1$ .

For any column  $j$  of  $\mathbf{A}_2$  there exists by hypothesis a convex combination  $\mathbf{y}^0$  of columns in  $\mathbf{A}_1$  which is strictly dominated by  $j$ , where we extend  $\mathbf{y}^0$  to be a strategy by using 0's for the columns of  $\mathbf{A}_2$ . It follows that

$$(\mathbf{x}', \mathbf{A}\beta_j) > (\mathbf{x}', \mathbf{A}\mathbf{y}^0) = (\mathbf{x}'\mathbf{A}, \mathbf{y}^0) \geq v \text{ for } j \text{ in } \mathbf{A}_2, \quad (2.2.3)$$

where the last inequality is a consequence of (2.2.2). The inequalities (2.2.2) and (2.2.3) together contradict the meaning of  $v$ . Thus no row of  $\mathbf{A}_3$  can be employed with positive probability in an optimal strategy.

An analogous argument shows that the columns of  $\mathbf{A}_2$  are not relevant.

Although all the dominance criteria are stated above in terms of strict inequalities, it is possible to apply these same theorems to nonstrict dominance. This again leads to a reduced matrix whose optimal strategies are also optimal for the original game, although in this case the sets of optimal strategies for the reduced game and the original game will not necessarily coincide. Once the value of the game and a pair of optimal strategies are ascertained, however, it is easy to check the relevance to any optimal strategy of the rows and columns that were eliminated.

It is possible to apply several of the above theorems successively to an initial matrix and the ensuing reduced matrices. Consider the following example:

$$\mathbf{A} = \begin{vmatrix} 0 & -3 & -3 & 5 & -3 \\ 1 & 3 & -2 & 2 & 0 \\ 1 & -2 & 1 & 1 & 0 \\ 0 & 0 & -1 & -4 & 1 \\ 3 & -2 & -1 & 6 & -2 \end{vmatrix}$$

The first row is dominated termwise by the fifth row. In the resulting

reduced matrix, the first column dominates the convex combination of the last two columns weighted equally. Deleting the first column, we obtain

$$\begin{vmatrix} 3 & -2 & 2 & 0 \\ -2 & 1 & 1 & 0 \\ 0 & -1 & -4 & 1 \\ -2 & -1 & 6 & -2 \end{vmatrix} = \begin{vmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{vmatrix},$$

where each of the  $\mathbf{A}_i$  represents the corresponding  $2 \times 2$  submatrix. Applying Theorem 2.2.4, we obtain the  $2 \times 2$  game

$$\begin{vmatrix} 3 & -2 \\ -2 & 1 \end{vmatrix},$$

whose solution for both players is the strategy  $\langle \frac{3}{8}, \frac{5}{8} \rangle$ . The value of the game is  $-\frac{1}{8}$ . It is readily verified that the only equalizing rows of  $\mathbf{A}$  with respect to the optimal strategy  $\langle 0, \frac{3}{8}, \frac{5}{8}, 0, 0 \rangle$  are the second and third rows, and that the same is true for the columns. We therefore conclude that this single strategy  $\langle 0, \frac{3}{8}, \frac{5}{8}, 0, 0 \rangle$  is the unique optimal strategy for each player.

**2.3 Construction of optimal strategies.** It was proved in Chapter 1 that matrix games have a value and that each player has an optimal strategy. However, these proofs present no effective method for finding the value of the game or determining explicit solutions. Before tackling the problem of characterizing solutions in general, let us consider some simple constructive methods of solving certain particularly important games. The large number of  $2 \times m$  and  $n \times 2$  games that arise in practice justifies consideration of the following methodology.

*Example 1.  $2 \times m$  game.* Consider the game in which Player I has 2 possible pure strategies, Player II has  $m$  possible pure strategies, and the pay-off matrix is

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \end{vmatrix}.$$

A strategy for Player I can be characterized by an  $x$  ( $0 \leq x \leq 1$ ), where  $x$  represents the probability of Player I's choosing strategy 1. If Player II chooses a pure strategy  $j$ , the pay-off corresponding to the strategies  $x$  can be represented by a line of the form  $l_j(x) = xa_{1j} + (1 - x)a_{2j}$ .

Let

$$\begin{aligned} \phi(x) &= \text{envelope from below of the set of lines } l_j(x) \\ &= \inf_j l_j(x) = \min_j l_j(x). \end{aligned}$$

Let  $x_0$  be such that  $\phi(x_0) = \max_x \phi(x)$ . Then  $\phi(x_0)$  is the maximum pay-off that Player I can assure himself by choosing  $x_0$  against any pure strategy that Player II might use.

Since a mixed strategy for Player II necessarily consists of a convex combination of the lines  $l_j(x)$ , it is clear that by choosing  $x_0$  Player I in all circumstances assures himself of at least the amount  $v^* = \phi(x_0)$ . Hence  $v \geq v^*$ , where  $v$  is the value of the game.

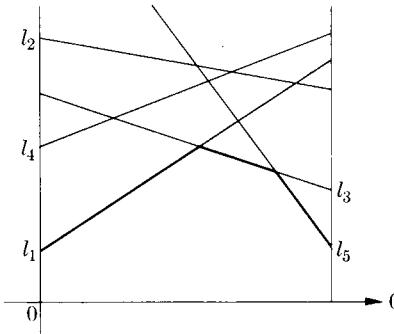


FIGURE 2.1

Now let us see to what amount Player II can restrict Player I. Let us assume that  $\phi(x)$  achieves a maximum at a single point  $x_0$  in the interior of the unit interval, as indicated in Fig. 2.1. If there exists a line  $l_{j_0}$  passing through  $\{x_0, \phi(x_0)\}$  with zero slope, then let the pure strategy  $j_0$  be II's strategy. If there exists no such line through  $\{x_0, \phi(x_0)\}$ , choose two lines,  $l_{j_0}$  and  $l_{j_0^*}$ , through the point  $\{x_0, \phi(x_0)\}$ , one with a positive slope and one with a negative slope. (The restriction that the minimum is attained at a single interior point obviously implies the existence of two such lines.) Also choose a weight  $\lambda_0$  ( $0 \leq \lambda_0 \leq 1$ ) such that the line  $l^* = \lambda_0 l_{j_0}(x) + (1 - \lambda_0) l_{j_0^*}(x)$  has zero slope, i.e.,

$$\lambda_0 l_{j_0}(x) + (1 - \lambda_0) l_{j_0^*}(x) \equiv v^* = \phi(x_0).$$

Let II's strategy be the mixed strategy

$$\langle 0, 0, \dots, \lambda_0, \dots, 1 - \lambda_0, \dots, 0 \rangle,$$

where  $\lambda_0$  occurs in the  $j_0$  coordinate and  $1 - \lambda_0$  in the  $j_0^*$  coordinate. If  $\phi(x)$  achieves its maximum at more than one point in the interior of the unit interval, it follows that there must exist a line  $l_{j_0}$  of the form  $l_{j_0}(x) \equiv v^*$ . Let II's strategy be the pure strategy  $j_0$  in this case. If the point  $x_0 = 0$  is a maximum point of  $\phi(x)$ , let II's strategy be the pure strategy corre-

sponding to the line passing through  $\{0, \phi(0)\}$  whose nonpositive slope has the largest absolute value; and if  $x_0 = 1$  is a maximum point, let II's strategy be the pure strategy corresponding to the line passing through  $\{1, \phi(1)\}$  with the largest nonnegative slope.

In all these cases, if II uses the designated strategy, I is limited to the amount  $v^*$  regardless of what strategy he plays. Therefore the value of the game must be less than or equal to  $v^*$ . But we found previously that  $v \geq v^*$ . Hence  $v = v^*$ , and the optimal strategies for I and II are those that were suggested as possible optimal strategies.

If the maximum point of  $\phi(x)$  occurs at one of the end points, both players have optimal pure strategies. If the function  $\phi(x)$  has a plateau, i.e., more than one maximum point, then II has an optimal pure strategy and I has an infinite number of optimal mixed strategies.

For an  $n \times 2$  game there are analogous procedures for finding the value and the optimal strategies. The lines  $l_i(y) = ya_{i1} + (1 - y)a_{i2}$ , where  $y$  varies over the unit interval, give the pay-off if I uses the pure strategy  $i$  and II the mixed strategy  $y$ . The minimum of the function  $\psi(y) = \sup_i l_i(y)$  gives the value of the game.

*Example 2.*  $3 \times m$  game. A strategy for Player I in a  $3 \times m$  game is designated by a pair  $\langle x_1, x_2 \rangle$ , where  $x_1 \geq 0$ ,  $x_2 \geq 0$ , and  $x_1 + x_2 \leq 1$ . The set of possible strategies can be represented as the shaded triangle in Fig. 2.2.

Another way to represent a strategy for Player I is by a plane triangle (Fig. 2.3). The three vertices correspond to the three pure strategies. Any other point  $\mathbf{x}$  in the triangle is a mixed strategy and is denoted by a triple  $\langle x_1, x_2, x_3 \rangle$ , where  $x_i$  is that part of a unit mass that would have to be located at end point  $i$  for the center of gravity of the system to be the point  $\mathbf{x}$ . Thus  $\langle x_1, x_2, x_3 \rangle$  designates  $\mathbf{x}$  in barycentric coordinates. The important feature of this use of barycentric coordinates to represent the set of mixed strategies is that it is independent of any coordinate system.

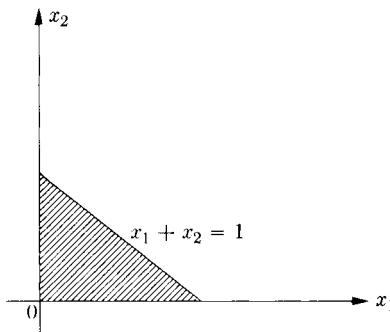


FIGURE 2.2

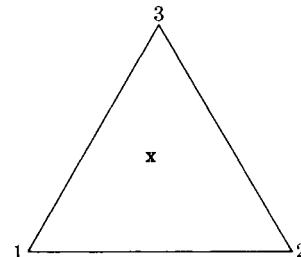


FIGURE 2.3

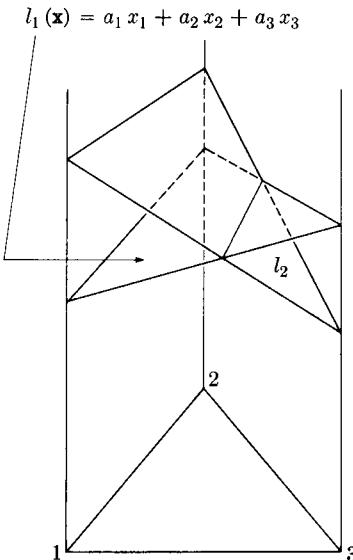


FIGURE 2.4

Over this plane triangle we can construct a set of planes, each plane giving the pay-off if Player II uses a particular pure strategy. This has been done in Fig. 2.4. As before, we form the envelope from below and consider its set of maximum points; this can be a single point, a line, or (in general) a polyhedron. The value of the function  $\phi(\mathbf{x}) = \inf_i l_i(\mathbf{x})$  at a maximum point is the value of the game. Points in the projection of the maximum polyhedron on the plane triangle correspond to optimal strategies for Player I. Optimal strategies for Player II are obtained by a correct weighting of the pure strategy planes in a manner analogous to the procedures used in the  $2 \times m$  game; i.e., the pure strategy planes are so weighted that the mixed strategy is a plane parallel to the basic triangle.

For  $n \times m$  games with both  $n$  and  $m$  greater than 3, this method becomes impractical.

**2.4 Characterization of extreme-point optimal strategies.** A strategy  $\mathbf{x}^0$  is called an extreme-point optimal strategy if  $\mathbf{x}^0$  is an extreme point of the convex set  $X^0$  of optimal strategies (see Section B.2). Extreme-point optimal  $\mathbf{y}$  strategies are similarly defined. By way of introduction to the methods used to characterize extreme-point optimal strategies, we shall first analyze a special case of Theorem 2.4.1.

Suppose that  $\mathbf{A} = [a_{ij}]$  is  $n \times n$  and nonsingular. Let  $\{\mathbf{x}^0, \mathbf{y}^0\}$  be a pair of optimal strategies and assume that  $x_i^0 > 0$  for all  $i$  and  $y_j^0 > 0$  for all  $j$ . Let  $\mathbf{e} = \langle 1, 1, \dots, 1 \rangle$  represent a vector which consists of  $n$  components each of value 1.

By Lemma 2.1.2

$$\mathbf{A}\mathbf{y}^0 \equiv \mathbf{v}, \quad \mathbf{x}^0\mathbf{A} \equiv \mathbf{v}, \quad (2.4.1)$$

where  $v$ , the value of the game, is to be interpreted either as a scalar or as a vector of suitable size all of whose components are  $v$ , depending on the context. It follows that

$$\mathbf{y}^0 = \mathbf{A}^{-1}\mathbf{v} = \mathbf{v}\mathbf{A}^{-1}$$

and

$$1 = \sum_{j=1}^n y_j^0 = (\mathbf{e}, \mathbf{A}^{-1}\mathbf{v}\mathbf{e}) = v(\mathbf{e}, \mathbf{A}^{-1}\mathbf{e}).$$

This last relation implies  $v \neq 0$ . Thus for this case the solution  $\mathbf{x}^0, \mathbf{y}^0$  and the value  $v$  can be explicitly computed by means of the following formulas:

$$v = \frac{1}{(\mathbf{e}, \mathbf{A}^{-1}\mathbf{e})}, \quad \mathbf{y}^0 = \frac{\mathbf{A}^{-1}\mathbf{e}}{(\mathbf{e}, \mathbf{A}^{-1}\mathbf{e})}, \quad \mathbf{x}^0 = \frac{\mathbf{e}\mathbf{A}^{-1}}{(\mathbf{e}, \mathbf{A}^{-1}\mathbf{e})}. \quad (2.4.2)$$

Clearly, a pair of solutions can be represented by the formulas of (2.4.2) whenever (2.4.1) holds and  $\mathbf{A}$  is nonsingular. The assumption that (2.4.1) is satisfied is in fact weaker than the hypothesis that  $x_i^0 > 0$  for all  $i$  and  $y_j^0 > 0$  for all  $j$ .

This example indicates that under suitable conditions, when  $\mathbf{A}$  is nonsingular, a pair of solutions of the game, one for each player, can be calculated explicitly in terms of  $\mathbf{A}^{-1}$ . To generalize this construction, we shall determine a submatrix  $\mathbf{M}$  of  $\mathbf{A}$  with properties similar to those of the matrix of (2.4.2). The optimal strategies obtained from such submatrices constitute extreme-point solutions. Our next theorem makes explicit this extension of (2.4.2).

► **THEOREM 2.4.1.** If  $\mathbf{x}^0, \mathbf{y}^0$  are extreme-point optimal strategies for a game with pay-off matrix  $\mathbf{A}$ , and if  $v \neq 0$ , then there exists a nonsingular submatrix  $\mathbf{M}$  of  $\mathbf{A}$  such that

$$v = \frac{1}{(\mathbf{e}, \mathbf{M}^{-1}\mathbf{e})} \quad (2.4.3a)$$

$$\mathbf{x}^0 = \frac{\mathbf{e}\mathbf{M}^{-1}}{(\mathbf{e}, \mathbf{M}^{-1}\mathbf{e})} \quad (2.4.3b)$$

$$\mathbf{y}^0 = \frac{\mathbf{M}^{-1}\mathbf{e}}{(\mathbf{e}, \mathbf{M}^{-1}\mathbf{e})}, \quad (2.4.3c)$$

where  $\mathbf{e}$  is a vector of the same dimension as  $\mathbf{M}$  all of whose components are 1.

(Formulas (2.4.3b) and (2.4.3c) exhibit only such components of  $\mathbf{x}^0, \mathbf{y}^0$  as correspond to the rows and columns of  $\mathbf{A}$  which appear in  $\mathbf{M}$ . The remaining components of  $\mathbf{x}^0, \mathbf{y}^0$  are identically zero.)

For convenience, the proof of this theorem is divided into a series of lemmas and discussions.

By  $X^0, Y^0$  we denote the spaces of optimal strategies for Players I and II, respectively. We know that  $X^0, Y^0$  are closed convex sets (Lemma 2.1.1). By Lemma B.2.4, we know that  $X^0, Y^0$  have extreme points. Let  $\mathbf{x}^0$  and  $\mathbf{y}^0$  be such points.

We can relabel the rows and columns of  $\mathbf{A} = \{a_{ij}\}$  in such a way that the first  $r$  components of  $\mathbf{x}^0$  and the first  $s$  components of  $\mathbf{y}^0$  are positive, while the other components are zero. Thus

$$\begin{aligned}\mathbf{x}^0 &= \langle x_1^0, \dots, x_r^0, 0, \dots, 0 \rangle \quad r \leq n, \\ \mathbf{y}^0 &= \langle y_1^0, \dots, y_s^0, 0, \dots, 0 \rangle \quad s \leq m.\end{aligned}$$

By Lemma 2.1.2, since  $\mathbf{x}^0$  is optimal and  $y_j^0 > 0$ , the inner product of  $\mathbf{x}^0$  with the  $j$ th column of  $\mathbf{A}$  equals  $v$  for all  $j \leq s$ . For the other columns the inner product is greater than or equal to  $v$ .

Similarly, the inner product of  $\mathbf{y}^0$  with the  $i$ th row equals  $v$  for all  $i \leq r$ . For the other rows, the inner product is less than or equal to  $v$ .

Again, we can rearrange the rows and columns of matrix  $\mathbf{A}$  (while maintaining the first  $r$  rows and  $s$  columns) so that (a) the inner product of  $\mathbf{y}^0$  and  $\mathbf{x}^0$  with each of the first  $\bar{r}$  rows and  $\bar{s}$  columns, respectively, is exactly equal to  $v$ , where  $0 < r \leq \bar{r} \leq n$ ,  $0 < s \leq \bar{s} \leq m$ ; (b) the inner product of  $\mathbf{y}^0$  with each of the last  $n - \bar{r}$  rows is less than  $v$ ; and (c) the inner product of  $\mathbf{x}^0$  with each of the last  $m - \bar{s}$  columns is greater than  $v$ .

With the rows and columns arranged in the order indicated, we decompose  $\mathbf{A}$  as

$$\mathbf{A} = \begin{vmatrix} \mathbf{A}_1 & \mathbf{A}_3 & \mathbf{A}_7 \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_8 \\ \mathbf{A}_5 & \mathbf{A}_6 & \mathbf{A}_9 \end{vmatrix}$$

where  $\mathbf{A}_1$  is of size  $r \times s$ ,  $\mathbf{A}_2$  of size  $(\bar{r} - r) \times s$ ,  $\mathbf{A}_3$  of size  $r \times (\bar{s} - s)$ ,  $\mathbf{A}_4$  of size  $(\bar{r} - r) \times (\bar{s} - s)$ , etc. This notation is used in the following lemma.

► **LEMMA 2.4.1.** The matrix

$$\mathbf{C} = \begin{vmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{vmatrix}$$

defines a linear mapping of  $E^s$  into  $E^{\bar{r}}$  with the property that only the zero vector is mapped by  $\mathbf{C}$  into zero.

*Proof* (by contradiction). We shall show that if  $\mathbf{C}\mathbf{z} = \mathbf{0}$  for  $\mathbf{z} \neq \mathbf{0}$ , then  $\mathbf{y}^0$  is not an extreme point of  $Y^0$ . First, note that  $\mathbf{C}\mathbf{z} = \mathbf{0}$  implies

$$0 = (\mathbf{x}^0, \mathbf{A}\mathbf{z}) = (\mathbf{x}^0\mathbf{A}, \mathbf{z}) = v \sum_{i=1}^s z_i.$$

But, since  $v \neq 0$ ,

$$\sum_{i=1}^s z_i = 0.$$

Next extend  $\mathbf{z}$  to the  $m$ -dimensional vector

$$\tilde{\mathbf{z}} = \langle z_1, z_2, \dots, z_s, 0, 0, \dots, 0 \rangle$$

and choose  $\epsilon$  so small that  $y_i^0 \pm \epsilon z_i \geq 0$  for all  $i$ . Then  $\mathbf{y}^0 \pm \epsilon \tilde{\mathbf{z}}$  are strategies, since

$$\sum_{i=1}^m (y_i^0 \pm \epsilon z_i) = \sum_{i=1}^s y_i^0 \pm \epsilon \sum_{i=1}^m z_i = 1.$$

Also,

$$\sum_{j=1}^s a_{ij}(y_j^0 \pm \epsilon z_j) = \sum_{j=1}^s a_{ij}y_j^0 \pm \epsilon \sum_{j=1}^s a_{ij}z_j = \begin{cases} v & i \leq \bar{r}, \\ v - \lambda_i \pm \epsilon K_i & i > \bar{r}, \end{cases}$$

where  $\lambda_i > 0$ , since

$$\sum_{j=1}^s a_{ij}y_j^0 < v$$

for all  $i > \bar{r}$ . Since we may choose  $\epsilon$  even smaller if necessary, it follows that  $\mathbf{y}^0 \pm \epsilon \tilde{\mathbf{z}}$  are optimal strategies. Finally, the relation

$$\mathbf{y}^0 = \frac{1}{2}(\mathbf{y}^0 + \epsilon \tilde{\mathbf{z}}) + \frac{1}{2}(\mathbf{y}^0 - \epsilon \tilde{\mathbf{z}})$$

shows that  $\mathbf{y}^0$  is not extreme. From this contradiction we infer the truth of the lemma.

Next we define  $\tilde{\mathbf{C}}$  to be a minimal submatrix of  $\mathbf{C}$  such that (a)  $\tilde{\mathbf{C}}$  contains the rows and columns of  $\mathbf{A}_1$ , and (b)  $\tilde{\mathbf{C}}\mathbf{z} = \mathbf{0}$  implies  $\mathbf{z} = \mathbf{0}$ . We know from Lemma 2.4.1 that  $\tilde{\mathbf{C}}$  exists.

The minimality of  $\tilde{\mathbf{C}}$  means that if  $\mathbf{C}^*$  is a proper submatrix of  $\tilde{\mathbf{C}}$  containing  $\mathbf{A}_1$ , condition (b) is violated. We shall use this fact later.

Analogously, we establish the existence of a minimal submatrix  $\tilde{\mathbf{D}}$  of  $\mathbf{D} = \|\mathbf{A}_1, \mathbf{A}_3\|$  such that (a')  $\tilde{\mathbf{D}}$  contains the rows and columns of  $\mathbf{A}_1$ , and (b')  $\mathbf{w}\tilde{\mathbf{D}} = \mathbf{0}$  implies  $\mathbf{w} = \mathbf{0}$ . We are now prepared to construct the submatrix  $\mathbf{M}$  needed in Theorem 2.4.1.

Let  $\mathbf{M}$  be the smallest submatrix of  $\mathbf{A}$  containing  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{D}}$ ; that is,  $\mathbf{M}$  consists of the rows represented in  $\tilde{\mathbf{C}}$  and the columns represented in  $\tilde{\mathbf{D}}$ . Clearly,  $\mathbf{M}$  is a matrix of size  $r' \times s'$ , where  $r \leq r' \leq \bar{r}$  and  $s \leq s' \leq \bar{s}$ .  $\mathbf{M}$  is called a kernel of the pair of extreme-point solutions  $\{\mathbf{x}^0, \mathbf{y}^0\}$ . Although  $\mathbf{M}$  is not uniquely determined, since  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{D}}$  were not uniquely defined, in most practical cases  $\mathbf{M}$  is unique for a specified  $\{\mathbf{x}^0, \mathbf{y}^0\}$ .

►LEMMA 2.4.2. The matrix  $\mathbf{M}$  is square and nonsingular.

*Proof.* We shall prove that  $\mathbf{M}\mathbf{z} = \mathbf{0}$  implies  $\mathbf{z} = \mathbf{0}$ . A similar argument shows that  $\mathbf{w}\mathbf{M} = \mathbf{0}$  implies  $\mathbf{w} = \mathbf{0}$ . These two results imply the validity of the lemma.

Let  $k$  be a fixed index such that  $s + 1 \leq k \leq s'$ . If  $s = s'$ , the argument of the succeeding paragraph is unnecessary. We wish first to show that  $\mathbf{M}\mathbf{z} = \mathbf{0}$  implies  $z_k = 0$  (the  $k$ th component of  $\mathbf{z}$  is zero).

Let  $\mathbf{D}^*$  be the submatrix of  $\tilde{\mathbf{D}}$  obtained by deleting the  $k$ th column. Since  $\tilde{\mathbf{D}}$  is minimal, there exists a nonzero vector  $\mathbf{w}^0$  of the form  $\mathbf{w}^0 = \langle w_1^0, w_2^0, \dots, w_r^0, 0, 0, \dots, 0 \rangle$  such that  $\mathbf{w}^0\mathbf{D}^* = \mathbf{0}$  and  $\mathbf{w}^0\tilde{\mathbf{D}}$  is nonzero in the  $k$ th coordinate only. Equivalently, the  $k$ th coordinate  $a_k$  of  $\mathbf{w}^0\mathbf{M} = \mathbf{w}^0\tilde{\mathbf{D}}$  is nonzero, while all other coordinates are zero. Hence  $a_k z_k = (\mathbf{w}^0\mathbf{M}, \mathbf{z}) = (\mathbf{w}^0, \mathbf{M}\mathbf{z}) = 0$ ; and since  $a_k \neq 0$ , we deduce that  $z_k = 0$ .

To sum up, if  $\mathbf{M}\mathbf{z} = \mathbf{0}$ , then  $\mathbf{z}$  necessarily has the form

$$\mathbf{z} = \langle z_1, z_2, \dots, z_s, 0, 0, \dots, 0 \rangle.$$

But then  $\mathbf{M}\mathbf{z} = \mathbf{0}$  is the same as  $\tilde{\mathbf{C}}\mathbf{z} = \mathbf{0}$ , and  $\mathbf{z}$  must be identically zero by Lemma 2.4.1. The proof is complete.

*Proof of Theorem 2.4.1.* Let  $\mathbf{M}$  denote the kernel associated with the extreme-point optimal strategies  $\mathbf{x}^0$  and  $\mathbf{y}^0$  as defined above. Since  $\mathbf{M}$  is a submatrix of

$$\begin{vmatrix} \mathbf{A}_1 & \mathbf{A}_3 \\ \mathbf{A}_2 & \mathbf{A}_4 \end{vmatrix},$$

we have

$$\mathbf{M}\dot{\mathbf{y}}^0 = \mathbf{v}, \quad \dot{\mathbf{x}}^0\mathbf{M} = \mathbf{v}, \quad (2.4.4)$$

where  $\mathbf{v}$  is again a vector of the proper size with each component equal to  $v$ , and  $\dot{\mathbf{y}}^0$  and  $\dot{\mathbf{x}}^0$  are used for the vectors  $\mathbf{y}^0$  and  $\mathbf{x}^0$  with some of the zero components deleted. But  $\mathbf{M}$  is nonsingular by Lemma 2.4.2, and therefore the equations of (2.4.4) become equivalent to formulas (2.4.3a, b, c).

The converse proposition to Theorem 2.4.1 takes the following form.

►THEOREM 2.4.2. If  $\mathbf{x}^0, \mathbf{y}^0$  are optimal strategies expressed by formulas (2.4.3a, b, c) and  $\mathbf{M}$  is a nonsingular submatrix of  $\mathbf{A}$ , then  $\mathbf{x}^0, \mathbf{y}^0$  are extreme-point optimal strategies.

*Proof.* Assume

$$\mathbf{x}^0 = \frac{\mathbf{x}' + \mathbf{x}''}{2},$$

with  $\mathbf{x}'$  and  $\mathbf{x}''$  optimal: Since the components of  $\mathbf{x}^0$  which do not correspond to rows of  $\mathbf{A}$  appearing in  $\mathbf{M}$  are identically zero, the corresponding components of  $\mathbf{x}'$  and  $\mathbf{x}''$  must also be zero. Thus it makes sense to speak about the vectors  $\mathbf{x}'\mathbf{M}$  and  $\mathbf{x}''\mathbf{M}$ .

Since  $\mathbf{x}'$  and  $\mathbf{x}''$  are optimal,

$$\mathbf{x}'\mathbf{M} \geq \mathbf{v} \quad (2.4.5)$$

and

$$\mathbf{x}''\mathbf{M} \geq \mathbf{v}. \quad (2.4.6)$$

Adding, we obtain  $\mathbf{x}^0\mathbf{M} \geq \mathbf{v}$ .

If strict inequality occurs in either (2.4.5) or (2.4.6) for some component, then correspondingly  $\mathbf{x}^0\mathbf{M} > \mathbf{v}$  for the same component. This is impossible by (2.4.3b). Therefore  $\mathbf{x}'\mathbf{M} = \mathbf{v}$  and  $\mathbf{x}''\mathbf{M} = \mathbf{v}$ ; and thus  $(\mathbf{x}' - \mathbf{x}'')\mathbf{M} = \mathbf{0}$ , which implies  $\mathbf{x}' = \mathbf{x}''$ , since  $\mathbf{M}$  is nonsingular. But if

$$\mathbf{x}^0 = \frac{\mathbf{x}' + \mathbf{x}''}{2},$$

with  $\mathbf{x}'$  and  $\mathbf{x}''$  optimal, always requires that  $\mathbf{x}' = \mathbf{x}''$ , then  $\mathbf{x}^0$  by definition is an extreme point of  $X^0$ . A similar argument applies to the strategy  $\mathbf{y}^0$ .

*The case  $v = 0$ .* If  $v = 0$ , we can choose any fixed  $a \neq 0$  and form the matrix  $\|a_{ij} + a\|$ . The value of the game is increased by  $a$  while the optimal strategies are unchanged (see Problem 7 of Chapter 1). Formulas (2.4.3a, b, c), are valid for the new game, whose value is now  $v = a$ . Our object now is to exhibit clearly the functional dependence of these formulas on the constant  $a$ . Let  $|\mathbf{M}|$  denote the determinant of  $\mathbf{M}$ . When  $v \neq 0$ , formulas (2.4.3a, b, c) can be written in terms of the adjoint matrix as follows:

$$v = \frac{|\mathbf{M}|}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})}, \quad (2.4.7a)$$

$$\mathbf{x}^0 = \frac{\mathbf{e} \text{ adj}(\mathbf{M})}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})}, \quad (2.4.7b)$$

$$\mathbf{y}^0 = \frac{\text{adj}(\mathbf{M})\mathbf{e}}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})}. \quad (2.4.7c)$$

If  $v = 0$ , and we consider the matrix  $\|a_{ij} + a\|$  with  $a > 0$ , then every extreme-point solution  $\{\mathbf{x}^0, \mathbf{y}^0\}$  admits the representation

$$a + v = \frac{|\mathbf{M} + \mathbf{a}|}{(\mathbf{e}, \text{adj}(\mathbf{M} + \mathbf{a})\mathbf{e})} = a + \frac{|\mathbf{M}|}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})},$$

$$\mathbf{x}^0 = \frac{\mathbf{e} \text{adj}(\mathbf{M} + \mathbf{a})}{(\mathbf{e}, \text{adj}(\mathbf{M} + \mathbf{a})\mathbf{e})} = \frac{\mathbf{e} \text{adj}(\mathbf{M})}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})},$$

$$\mathbf{y}^0 = \frac{\text{adj}(\mathbf{M} + \mathbf{a})\mathbf{e}}{(\mathbf{e}, \text{adj}(\mathbf{M} + \mathbf{a})\mathbf{e})} = \frac{\text{adj}(\mathbf{M})\mathbf{e}}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})},$$

provided  $(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})$  is not zero by virtue of the identities

$$(1) \quad |\mathbf{M} + \mathbf{a}| = |\mathbf{M}| + a(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e}),$$

$$(2) \quad \mathbf{e} \text{adj}(\mathbf{M} + \mathbf{a}) = \mathbf{e} \text{adj}(\mathbf{M}),$$

$$(3) \quad \text{adj}(\mathbf{M} + \mathbf{a})\mathbf{e} = \text{adj}(\mathbf{M})\mathbf{e}$$

(see Section A.9). Hence (2.4.7a, b, c) are valid also for the case  $v = 0$ .

Theorem 2.4.1 and its converse can now be restated in terms of the adjoints of square submatrices instead of in terms of inverses. With this new formulation (Theorem 2.4.3) we need no longer distinguish between the two cases  $v = 0$  and  $v \neq 0$ .

►THEOREM 2.4.3. A necessary and sufficient condition that the optimal strategies  $\mathbf{x}^0$  and  $\mathbf{y}^0$  be extreme-point optimal strategies and  $v$  the value of the game is that there exist a square submatrix  $\mathbf{M}$  of  $\mathbf{A}$  such that

$$(1) \quad (\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e}) \neq 0, \quad (3) \quad \dot{\mathbf{x}}^0 = \frac{\mathbf{e} \text{adj}(\mathbf{M})}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})}, \quad (2.4.8)$$

$$(2) \quad v = \frac{|\mathbf{M}|}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})}, \quad (4) \quad \dot{\mathbf{y}}^0 = \frac{\text{adj}(\mathbf{M})\mathbf{e}}{(\mathbf{e}, \text{adj}(\mathbf{M})\mathbf{e})},$$

where  $\dot{\mathbf{x}}^0$  is obtained from  $\mathbf{x}^0$  by deleting the components corresponding to the rows deleted from  $\mathbf{A}$  to obtain  $\mathbf{M}$  (the deleted components equaling zero), and  $\dot{\mathbf{y}}^0$  is obtained from  $\mathbf{y}^0$  analogously.

Theorem 2.4.1 suggests the following systematic method of obtaining the value and extreme-point optimal strategies of a game. Let us restrict our attention to games with value different from zero. Take all possible submatrices  $\mathbf{M}$  of  $\mathbf{A}$ . For each  $\mathbf{M}$  check to see if it has an inverse. If no inverse exists, go on to another  $\mathbf{M}$ . If one does exist, compute  $\dot{\mathbf{x}}^0$ ,  $\dot{\mathbf{y}}^0$ , and  $v$  by formulas (2.4.3a, b, c). The relations  $\sum_i \dot{x}_i^0 = 1$  and  $\sum_j \dot{y}_j^0 = 1$  auto-

matically hold by virtue of (2.4.3c), but it might happen that  $x_i^0 < 0$  or  $y_j^0 < 0$  for some  $i$  or  $j$ . In this event, reject  $\mathbf{M}$ . If all the terms are non-negative, form  $\mathbf{x}^0$  and  $\mathbf{y}^0$  by adding zeros at the coordinates of the deleted rows and columns, respectively. From (2.4.3a) and (2.4.3b) it follows that the inner product of  $\mathbf{x}^0$  with any nondeleted column equals  $v$ , but a check must be made to see that the inner product of  $\mathbf{x}^0$  with a deleted column is greater than or equal to  $v$ . Similarly, the inner product of  $\mathbf{y}^0$  with any non-deleted row equals  $v$ , but a check must be made to see that the inner product of  $\mathbf{y}^0$  with a deleted row is less than or equal to  $v$ . If all these inner products correctly satisfy the conditions,  $\mathbf{x}^0$  and  $\mathbf{y}^0$  are extreme-point optimal strategies. If they do not, this  $\mathbf{M}$  must be rejected.

When the entire set of submatrices has been processed, we shall have the sets of extreme-point optimal strategies for Players I and II. The sets of optimal strategies are then obtained by forming the convex hull of these extreme points.

It should be emphasized that the class of all possible submatrices of  $\mathbf{A}$  includes  $\mathbf{A}$  itself and all single elements.

As an illustration of Theorem 2.4.3, consider the pay-off matrix

$$\begin{matrix} & & & \text{II} \\ & & & \\ \mathbf{I} & \left\| \begin{matrix} 1 & 1 & 3 \\ 1 & 1 & 0 \\ 0 & 2 & -5 \end{matrix} \right\| & = & \mathbf{A}. \end{matrix}$$

The four pairs of extreme-point optimal strategies and their corresponding kernels  $\mathbf{M}$  are as follows:

$$\mathbf{x}^0 = \langle 1, 0, 0 \rangle \quad \mathbf{y}^0 = \langle 1, 0, 0 \rangle \quad \mathbf{M} = a_{11} = 1$$

$$\mathbf{x}^0 = \langle 1, 0, 0 \rangle \quad \mathbf{y}^0 = \langle \frac{1}{2}, \frac{1}{2}, 0 \rangle \quad \mathbf{M} = \left\| \begin{matrix} 1 & 1 \\ 0 & 2 \end{matrix} \right\|$$

$$\mathbf{x}^0 = \langle \frac{1}{3}, \frac{2}{3}, 0 \rangle \quad \mathbf{y}^0 = \langle 1, 0, 0 \rangle \quad \mathbf{M} = \left\| \begin{matrix} 1 & 3 \\ 1 & 0 \end{matrix} \right\|$$

$$\mathbf{x}^0 = \langle \frac{1}{3}, \frac{2}{3}, 0 \rangle \quad \mathbf{y}^0 = \langle \frac{1}{2}, \frac{1}{2}, 0 \rangle \quad \mathbf{M} = \mathbf{A}.$$

**2.5 Completely mixed matrix games.** A matrix game is said to be completely mixed (c.m.) if every optimal  $\mathbf{x}, \mathbf{y}$  involves every pure strategy with positive probability: i.e., if  $\mathbf{x}^0 = \langle x_1^0, \dots, x_n^0 \rangle$  and  $\mathbf{y}^0 = \langle y_1^0, \dots, y_m^0 \rangle$  are both optimal, then  $x_i^0 > 0$  for all  $i$  and  $y_j^0 > 0$  for all  $j$ . Completely mixed matrix games constitute an important subclass of games for which the solutions are easily determined.

► **THEOREM 2.5.1.** The solutions to a game with matrix  $\mathbf{A}$  c.m. are unique.

The assertion follows from Theorem 2.4.3. If  $\mathbf{x}^0, \mathbf{y}^0$  are extreme-point optimal strategies, their kernel must be the full matrix  $\mathbf{A}$ . From (2.4.8) it is clear that all extreme-point optimal strategies coincide, since they share a common kernel. Hence the solutions for each player are unique.

► **COROLLARY 2.5.1.** If a game  $\mathbf{A}$  is c.m. and  $v \neq 0$ , then  $\mathbf{A}$  is nonsingular.

This follows from Theorem 2.4.1, since the kernel of any pair of extreme-point solutions must be  $\mathbf{A}$ .

*Example 1. Minkowski-Leontief matrix.* This is a matrix which occurs in economic theory in the area of activity analysis. A matrix  $\mathbf{A}$  is called a Minkowski-Leontief matrix if (a)  $\mathbf{A}$  is square; (b)  $a_{ii} > q \geq 0$ ; (c)  $a_{ij} \leq q$  for  $i \neq j$ ; and

$$(d) \quad \sum_{i=1}^n a_{ij} > nq \geq 0.$$

Usually in practice  $q = 0$ . We shall now show that a Minkowski-Leontief matrix generates a c.m. game.

Let  $\mathbf{x} = \langle 1/n, 1/n, \dots, 1/n \rangle$ ; then

$$\sum_{i=1}^n a_{ij}x_i = \frac{1}{n} \sum_{i=1}^n a_{ij} > q.$$

Consequently, the value of the game is  $v > q \geq 0$ . Suppose that  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  is an optimal strategy, with  $x_{j_0} = 0$ . Then

$$(\mathbf{x}\mathbf{A})_{j_0} = \sum_{i=1}^n x_i a_{ij_0} = \sum_{i \neq j_0} x_i a_{ij_0} \leq \sum_{i=1}^n x_i q \leq q < v.$$

Hence  $\mathbf{x}$  cannot be optimal. Thus every optimal strategy for Player I uses every component with positive probability.

Suppose that  $\mathbf{y} = \langle y_1, \dots, y_n \rangle$  is an optimal strategy for Player II and  $y_{j_0} = 0$ . By Lemma 2.1.2,

$$\sum_{j=1}^n a_{ij}y_j = v$$

for all  $i$ . However, for  $i = j_0$ ,

$$\sum_{j=1}^n a_{j_0j}y_j = \sum_{j \neq j_0} a_{j_0j}y_j \leq q < v,$$

a contradiction. Thus, every optimal  $\mathbf{y}$  strategy has every component positive. As a corollary, it follows that every Minkowski-Leontief matrix is nonsingular.

As an immediate consequence of Lemma 2.5.1 below and the above example, we find that if  $\mathbf{A}$  is a square matrix such that (a)  $a_{ij} \leq q$  for  $i \neq j$ , (b)  $a_{ii} > q \geq 0$ , and

$$(c) \quad \sum_{j=1}^n a_{ij} > nq,$$

then  $\mathbf{A}$  is c.m.

►LEMMA 2.5.1. If  $\mathbf{A}$  is completely mixed, then the matrix  $\mathbf{B} = \mathbf{A}'$  ( $\mathbf{A}$  transpose) is completely mixed, and conversely.

*Proof.* Let  $\mathbf{w}^0, \mathbf{z}^0$  be the unique solutions of  $\mathbf{A}$  for Players I and II, respectively. Then, by Lemma 2.1.2,

$$\mathbf{w}^0 \mathbf{A} = \mathbf{v}, \quad \mathbf{A} \mathbf{z}^0 = \mathbf{v}.$$

Let  $\mathbf{B} = \mathbf{A}'$ ; then  $\mathbf{B}' = (\mathbf{A}')' = \mathbf{A}$  and

$$\mathbf{z}^0 \mathbf{B} = \mathbf{A} \mathbf{z}^0 = \mathbf{v}, \quad \mathbf{B} \mathbf{w}^0 = \mathbf{w}^0 \mathbf{A} = \mathbf{v}.$$

Hence  $\mathbf{z}^0$  and  $\mathbf{w}^0$  comprise optimal strategies for the matrix  $\mathbf{B}$  for Players I and II, respectively, with the same value  $v$ . They involve all components with positive probability, and they must be unique or the solution of  $\mathbf{A}$  would not be unique by reversing the reasoning. Hence  $\mathbf{A}'$  is completely mixed.

The converse follows from the fact that  $(\mathbf{A}')' = \mathbf{A}$ .

*Example 2. Ordered set game.* Let  $L_0, L_1, \dots, L_{n-1}$  be disjoint closed sets on the real axis with  $L_i < L_{i+1}$ ; that is, the set  $L_i$  lies strictly to the left of the set  $L_{i+1}$ . Let

$$\mathbf{A} = \begin{vmatrix} L_0 & L_1 & L_2 & L_3 & \cdots & L_{n-1} \\ L_{n-1} & L_0 & L_1 & L_2 & \cdots & \cdots \\ L_{n-2} & L_{n-1} & L_0 & L_1 & \cdots & \cdots \\ \vdots & & & & & \\ L_1 & L_2 & \cdots & \cdots & \cdots & L_0 \end{vmatrix}$$

$\mathbf{A}$  is interpreted as a matrix  $\|a_{ij}\|$ , where the element  $a_{ij}$  is a value contained in the  $L$ -set appearing in the  $\{i, j\}$  place. Algebraically,  $a_{i,j} \in L_{j-i}$  if  $j \geq i$ , and  $a_{i,j} \in L_{n-i+j}$  if  $j < i$ .

We assert that  $\mathbf{A}$  is completely mixed. To show this, we assume that  $\mathbf{x}$  is optimal and that  $x_j = 0$  for some  $j$ . If  $i \neq j$ , then  $a_{ij} > a_{i,j-1}$  (if  $j = 0$ , then  $j - 1$  is interpreted as  $n - 1$ ). Since  $x_j = 0$ ,

$$\sum_{i=1}^n x_i a_{ij} > \sum_{i=1}^n x_i a_{i,j-1} \geq v.$$

Therefore, by Lemma 2.1.2, no optimal  $\mathbf{y}$  strategy can have  $y_j > 0$ . Clearly,  $a_{j+1,k} < a_{j,k}$  if  $j \neq k$  (interpret  $j+1 = 0$  if  $j = n-1$ ); hence if  $\mathbf{y}$  is optimal, we obtain

$$\sum_{k=1}^n a_{j+1,k} y_k < \sum_{k=1}^n a_{jk} y_k \leq v.$$

Again by Lemma 2.1.2, we see that no optimal  $\mathbf{x}$  strategy can have  $x_{j+1} > 0$ . By iteration, we obtain  $\mathbf{x} = \mathbf{0}$ , which is impossible; therefore every optimal  $\mathbf{x}$  has all components positive. Since starting with  $\mathbf{y}$  optimal and  $y_j = 0$  leads to the same contradiction, we see that every optimal  $\mathbf{y}$  also has all components positive. Hence  $\mathbf{A}$  is completely mixed.

If the sets  $L_k$  are ordered in the reverse direction, i.e.,

$$L_0 > L_1 > L_2 > \dots,$$

then the corresponding game is completely mixed.

In particular, if the principal diagonal of a  $2 \times 2$  matrix dominates or is dominated, then the optimal strategies for that matrix are completely mixed. Formally, the game with matrix

$$\mathbf{A} = \begin{vmatrix} a & c \\ d & b \end{vmatrix}$$

will be completely mixed provided

$$\max \{a, b\} < \min \{c, d\} \quad \text{or} \quad \min \{a, b\} > \max \{c, d\}.$$

The converse is also valid in this case, for if the principal diagonal does not dominate, a saddle point exists.

No such complete description of completely mixed games exists for games of higher order than  $2 \times 2$  (see Problem 9).

*Example 3. Cyclic matrix.* Let

$$\mathbf{A} = \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & \cdots \\ \vdots & & & & \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{vmatrix}$$

where  $a_i \neq a_j$ . This example is similar to the preceding one, with  $L_k = \{a_k\}$  consisting of a single point; however, we do not assume here that  $a_i < a_{i+1}$ , or, more generally, that any ordering relationship is imposed on the quantities  $a_i$ .

A sufficient condition that  $\mathbf{A}$  be c.m. is that  $\det \mathbf{A} \neq 0$ . In order to verify this statement, we first consider the form of  $\det \mathbf{A}$ .

Note that  $\lambda_r = \sum_{k=0}^{n-1} a_k w^{kr}$  ( $r = 1, \dots, n$ )

are distinct characteristic roots of  $\mathbf{A}$ , where  $w$  is a primitive  $n$ th root of unity. Thus

$$\det \mathbf{A} = \prod_{r=1}^n \lambda_r = \prod_{r=1}^n \sum_{k=0}^{n-1} a_k w^{kr}.$$

Next let us verify that  $\mathbf{x} = \mathbf{y} = \langle 1/n, \dots, 1/n \rangle$  are optimal strategies and

$$v = \frac{1}{n} \sum a_i.$$

Explicitly,

$$(\mathbf{x}\mathbf{A})_j = \sum_{i=1}^n x_i a_{ij} = \frac{1}{n} \sum_{i=1}^n a_i = v$$

and similarly  $\mathbf{A}\mathbf{y} = \mathbf{v}$ . Since, by assumption,  $\det \mathbf{A} \neq 0$ , it follows from Theorem 2.4.2 that  $\mathbf{x}^0 = \mathbf{y}^0 = \langle 1/n, \dots, 1/n \rangle$  are both extreme-point optimal strategies. Let  $\mathbf{y}'$  be any extreme-point optimal strategy different from  $\mathbf{y}^0$ . The pair of extreme-point solutions  $\{\mathbf{x}^0, \mathbf{y}'\}$  must have as its kernel the full matrix, and thus  $\mathbf{y}' = \mathbf{y}^0$ . In a similar manner we find that  $\mathbf{x}^0$  is the only optimal strategy for Player I. Hence  $\mathbf{A}$  is c.m.

In particular, for  $n = 3$ , where  $a_0, a_1, a_2$  are not all equal and  $a_0 + a_1 + a_2 \neq 0$ , we find that

$$\mathbf{A} = \begin{vmatrix} a_0 & a_1 & a_2 \\ a_2 & a_0 & a_1 \\ a_1 & a_2 & a_0 \end{vmatrix}$$

is completely mixed. It is enough to verify that  $a_0 = a_1 = a_2$  is implied by either  $a_0 + a_1w + a_2w^2 = 0$  or  $a_0 + a_1w^2 + a_2w = 0$ , where  $w$  is a third root of unity. By examining the imaginary part of either equation, we find that  $a_1 = a_2$ . It follows from the real part that  $a_0 = a_1 = a_2$ .

In contrast,

$$\begin{vmatrix} 0 & 1 & 3 & 2 \\ 2 & 0 & 1 & 3 \\ 3 & 2 & 0 & 1 \\ 1 & 3 & 2 & 0 \end{vmatrix}$$

has a value of  $\frac{3}{2}$ , and the set of extreme-point optimal strategies for Player I consists of the two strategies  $\langle \frac{1}{2}, 0, \frac{1}{2}, 0 \rangle$  and  $\langle 0, \frac{1}{2}, 0, \frac{1}{2} \rangle$ . The same is true for Player II.

**2.6 Symmetric games.** A symmetric game is a game with a skew-symmetric pay-off matrix ( $\mathbf{A}' = -\mathbf{A}$ ). The two players have available the same sets of strategies, and both can achieve the same return by employing a given strategy. Such a game is in all senses the fairest to both players because any opportunity open to either player is also open to the other. In fact, if Player I employs the strategy  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$ , his yield is the component of the vectors in  $\mathbf{x}\mathbf{A}$  that corresponds to the pure strategy used by Player II. Similarly, if Player II uses  $\mathbf{y} = \langle y_1, y_2, \dots, y_m \rangle$ , the possible yields to Player I are  $\mathbf{A}\mathbf{y}$  depending upon which pure strategy he uses. If  $\mathbf{x} = \mathbf{y}$ , then

$$-\mathbf{A}\mathbf{y} = \mathbf{x}\mathbf{A}. \quad (2.6.1)$$

This relationship points up the distinguishing feature of symmetric games: namely, that any outcome achievable by Player I can also be achieved by Player II by using the same strategy.

Since  $(\mathbf{x}, \mathbf{A}\mathbf{y}) = 0$  whenever  $\mathbf{x} = \mathbf{y}$ , we infer that the value of the game is zero. Furthermore, it is clear from (2.6.1) that the sets of optimal strategies  $X^0$  and  $Y^0$  are identical.

These special properties lend some importance and interest to symmetric games. Their main significance, however, derives from the fact that any game can be associated with a symmetric game by an induced correspondence that relates the sets of optimal strategies of the given game with those of the associated symmetric game. Two methods have been proposed for symmetrizing a matrix game of order  $n \times m$ . The first involves converting the game to a larger symmetric game of order  $n + m + 1$ . The second associates with the original game a larger symmetric matrix having  $n \times m$  rows. We now describe these two symmetrizing procedures.

► **THEOREM 2.6.1.** If  $\mathbf{A}$  is  $n \times m$ , then there exists a symmetric game  $\mathbf{A}_1$  of order  $(m + n + 1) \times (m + n + 1)$  such that every optimal strategy  $\langle x_1, \dots, x_n, y_1, \dots, y_m, \lambda \rangle$  for the game  $\mathbf{A}_1$  has the property  $\sum x_i = \sum y_i = a > 0$ , and  $\mathbf{x}/a, \mathbf{y}/a$  are optimal for the game  $\mathbf{A}$  with value  $\lambda/a$ ; and conversely.

*Proof.* We may assume  $a > 0$  for the game  $\mathbf{A}$ . Let

$$\mathbf{A}_1 = \begin{vmatrix} 0 & \mathbf{A} & -1 \\ -\mathbf{A}' & 0 & 1 \\ 1 & -1 & 0 \end{vmatrix}$$

$\mathbf{A}_1$  is skew-symmetric of order  $n + m + 1$ . Let  $\langle x_1, \dots, x_n, y_1, \dots, y_m, \lambda \rangle$  be optimal for Player II in the game  $\mathbf{A}_1$ . Then, since the value of the game  $\mathbf{A}_1$  is zero,

$$\mathbf{A}\mathbf{y} - \lambda \leq 0, \quad -\mathbf{x}\mathbf{A} + \lambda \leq 0, \quad \sum x_i - \sum y_i \leq 0. \quad (2.6.2)$$

Note that  $\lambda > 0$ ; otherwise we would have  $\mathbf{Ay} \leq \mathbf{0}$ ,  $\mathbf{xA} \geq \mathbf{0}$  in (2.6.2), which contradicts our assumption that  $v > 0$  for the game  $\mathbf{A}$ . But  $\lambda > 0 \rightarrow \sum x_i = \sum y_i$  by Lemma 2.1.2. Finally,  $\lambda < 1$  or the second relation of (2.6.2) could not be satisfied. Hence

$$\sum y_i = \sum x_i = \frac{1 - \lambda}{2} > 0.$$

If we set

$$a = \frac{1 - \lambda}{2},$$

we obtain from (2.6.2) the inequality

$$\mathbf{A} \left( \frac{\mathbf{y}}{a} \right) \leq \frac{\lambda}{a} \leq \left( \frac{\mathbf{x}}{a} \right) \mathbf{A}.$$

Therefore  $\mathbf{x}/a, \mathbf{y}/a$  are optimal for  $\mathbf{A}$ , and the value of the game  $\mathbf{A}$  is  $\lambda/a$ .

Conversely, if  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  and  $\mathbf{y} = \langle y_1, \dots, y_m \rangle$  are optimal for  $\mathbf{A}$ , then (2.6.2) can be verified for the vector

$$\mathbf{w} = \left\langle \frac{\mathbf{x}}{2 + v}, \frac{\mathbf{y}}{2 + v}, \frac{v}{2 + v} \right\rangle.$$

Thus  $\mathbf{w}$  is optimal for the game  $\mathbf{A}_1$ .

Our second method of symmetrizing a game proceeds as follows: Visualize a player engaged in two games  $\mathbf{B}'$  and  $\mathbf{B}''$ , in which he maximizes and minimizes, respectively. The value of this game is the cumulative expected value obtained by playing  $\mathbf{B}'$  and  $\mathbf{B}''$ . Suppose that the indices for  $\mathbf{B}'$  are  $k', l'$  and the indices for  $\mathbf{B}''$  are  $k'', l''$ . Player I controls  $k', l''$  and Player II controls  $k'', l'$ ; hence a strategy  $i$  for Player I is a choice of  $k'$  and  $l''$ , and a strategy  $j$  for Player II is a choice of  $k''$  and  $l'$ . The expected pay-off is  $b_{k'l'} + b_{k''l''}$ . Suppose now that  $\mathbf{B}'' = -\mathbf{B}$  and  $\mathbf{B}' = \mathbf{B}$ ; then  $A_{ij} = b_{k'l'} - b_{k''l''}$  for  $i = (k', l'')$ ,  $j = (k'', l')$ . This game is symmetric, and hence there exists an optimal  $y_j$  with components  $y_{k''l'}$  satisfying  $y_{k''l'} \geq 0$ ,  $\sum y_{k''l'} = 1$ . Put

$$\xi_k = \sum_l y_{kl} \quad \text{and} \quad \eta_l = \sum_k y_{kl}.$$

Then  $\xi_k \geq 0$ ,  $\eta_l \geq 0$  and  $\sum \xi_k = \sum \eta_l = 1$ .

It is easily shown that  $\eta$  and  $\xi$  are optimal strategies for  $\mathbf{B}$ .

## 2.7 Problems.

1. Use the graphical method to solve the game with matrix

$$\begin{vmatrix} 1 & 0 & 4 & -1 \\ -1 & 1 & -2 & 5 \end{vmatrix}.$$

2. Describe a graphical method for solving  $m \times 2$  games.  
 3. Describe all the extreme-point optimal strategies for the game with matrix

$$\mathbf{A} = \begin{vmatrix} a_1 & & & & \\ & a_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & a_n \end{vmatrix}$$

where  $a_i > 0$  for all  $i$ .

4. Solve Problem 3 without the restriction  $a_i > 0$ .  
 5. Use Theorem 2.4.1 to solve the game with matrix

$$\begin{vmatrix} 0 & 5 & 6 & -4 \\ 3 & 9 & 9 & -6 \\ 3 & -1 & 0 & 2 \end{vmatrix}.$$

6. Use Theorem 2.4.3 to solve the game of Two-finger Morra, described in Section 1.2.

7. The game with matrix

$$\begin{vmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{vmatrix}$$

has a pair of optimal strategies whose kernel is the full matrix. Find these strategies.

8. The game described in Problem 4 of Chapter 1 has a pair of optimal strategies that use only the last five rows and the last five columns of the matrix. Find these strategies.

9. Show that the game with matrix

$$\begin{vmatrix} 4 & -3 & -2 \\ -3 & 4 & -2 \\ 0 & 0 & 1 \end{vmatrix}$$

is completely mixed, and find the optimal strategies.

10. Find all solutions to the game with matrix

$$\begin{array}{c|cccccc} & 4 & 3 & 3 & 2 & 2 & 6 \\ \hline 6 & 0 & 4 & 2 & 6 & 2 & \\ 0 & 7 & 3 & 6 & 2 & 2 & \end{array}$$

11. Let  $\mathbf{A}$  be an  $m \times n$  matrix with

$$a_{i,j} \leq a_{i,j+1} \quad (i = 1, \dots, m; j = 1, \dots, n-1).$$

Show that  $\mathbf{A}$  always has a saddle point.

12. A cat and a mouse simultaneously enter this blind maze at the points  $A$  and  $B$ , respectively; they may turn corners but they cannot turn around or stop. They travel at the same speed, and they have enough time to cover four line segments. If the cat finds the mouse, he wins one unit. Find the optimal strategies.

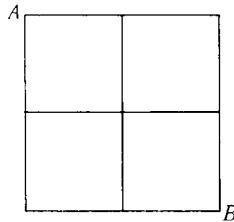


FIGURE 2.5

13. Use the dominance theorems to solve the game

$$\begin{array}{c|ccccc} & 1 & 1 & 0 & 0 & \\ & 1 & 1 & 1 & 0 & \circ \\ \hline 0 & 1 & 1 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & 1 & 1 & 1 & 0 \\ \hline \circ & & 0 & 1 & 1 & 1 & \\ & 0 & 0 & 1 & 1 & & \end{array}$$

14. If  $\mathbf{A}$  denotes a matrix all of whose elements are nonnegative, and every column of  $\mathbf{A}$  has at least one positive component, prove that the value of the game with matrix  $\mathbf{A}$  is positive.

15. Prove that the set of optimal strategies in a matrix game constitutes a polyhedral convex set.

16. Let  $a_1 < a_2 < a_3 < a_4 < a_5 < a_6$ , and

$$\mathbf{A} = \begin{vmatrix} a_4 & a_3 & a_3 \\ a_1 & a_6 & a_5 \\ a_2 & a_4 & a_3 \end{vmatrix}$$

Show that any optimal strategy satisfies the conditions

$$x_3 = y_2 = 0, \quad x_1 > \frac{1}{2},$$

$$x_1 > y_1 > x_2, \quad a_3 < v < a_4.$$

17. Discuss the relationship between the optimal strategies for the game  $\mathbf{A}$  and the optimal strategies for the symmetric game

$$\begin{vmatrix} 0 & \mathbf{A} & -1 & 1 \\ -\mathbf{A}' & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{vmatrix}$$

18. Let  $G^k$  ( $k = 1, \dots, K$ ) be a game with matrix  $\|G_{ij}^k\|$ , where  $i = 1, \dots, m$  and  $j = 1, \dots, n$ . Let  $X^k$  and  $v^k$  be the set of optimal maximizing strategies and the value of  $G^k$ , respectively. Show that if  $\cap X^k \neq 0$ , the value  $v$  of the game  $G$  with matrix

$$\|G_{ij}\| = \left\| \sum_{k=1}^K G_{ij}^k \right\|$$

satisfies the inequality

$$v \geq \sum_{k=1}^K v^k.$$

19. Consider a game whose matrix  $\mathbf{B}$  can be decomposed into  $MN$  sub-matrices  $\mathbf{B}_j^i$  with values  $v_j^i$ . Show that if for each  $i$  there is a common optimal strategy for the games  $\mathbf{B}_j^i$ , then the value of  $\mathbf{B}$  is greater than or equal to the value of the game  $v$  whose matrix is  $\|v_j^i\|$ . Describe the corresponding result in terms of the games  $\mathbf{B}_j^i$  (with fixed  $j$ ), and show that if both conditions hold, the value of  $\mathbf{B}$  and its optimal strategies can be determined from the values and optimal strategies of the games  $\mathbf{B}_j^i$  and  $v$ .

20. Show that there exists no symmetric completely mixed game of even order.

21. Let  $a_1, \dots, a_n$  be positive and  $b_1, \dots, b_{n-1}$  be negative numbers such that  $b_{i-1} + a_i + b_i \geq 0$  ( $i = 2, \dots, n - 1$ ),  $a_1 + b_1 \geq 0$ , and  $a_n + b_{n-1} \geq 0$ . Solve the game with matrix

$$\begin{array}{c|ccc|c} & a_1 & b_1 & & \\ \hline a_1 & & & & \\ b_1 & a_2 & b_2 & & \circ \\ b_2 & a_3 & b_3 & & \\ \vdots & \vdots & \vdots & & \\ & \vdots & \vdots & & \\ \hline & & & b_{n-1} & \\ \circ & & & & \\ b_{n-1} & & a_n & & \end{array}$$

22. Show that if

$$a_{i,j-1} - 2a_{i,j} + a_{i,j+1} \geq 0$$

for all  $i$  and  $j$ , then the game with matrix  $\mathbf{A}$  always has an optimal strategy in which each player uses at most two components.

## NOTES AND REFERENCES TO CHAPTER 2

**§ 2.1.** The simple properties possessed by the solutions of games described in this section were known and used by von Neumann. They are implicitly contained in [242].

**§ 2.2.** Dominance concepts were introduced first in [242]. The two forms of dominance embodied in Theorems 2.2.1 and 2.2.2 are standard equipment in the analysis of game problems. The earliest formulation of the four types of dominance cannot be associated with any individual; in general, it may be ascribed to the Rand Corporation, where many of the more extensive early mathematical developments of game theory were achieved. All four forms of dominance are mentioned in Dresher's handbook [66].

**§ 2.3.** The geometric method of this section is well known and has been duly elaborated upon in most elementary treatments of game theory, e.g., [179] and [169]. Borel [39] used these ideas in proving the min-max theorem for  $2 \times 2$  games.

**§ 2.4.** The principal theorem of this section was discovered by Snow and Shapley [219]. One can summarize the main idea of their proof as follows: A point in  $E^n$  space can be located as the unique intersection of  $n$  particular linearly independent hyperplanes. However, if  $k$  hyperplanes through a point are prescribed, some of which may be linearly dependent, then it is necessary to add at least  $n - k$  hyperplanes to ensure the point determined. The extra hyperplanes

beyond  $n - k$  are required by virtue of the dependence relations of the prescribed  $k$  hyperplanes.

An elaboration on the geometry underlying this theorem was converted by Motzkin, Raiffa, Thompson, and Thrall [187] into a systematic method for determining both the value and all the solutions of a two-person zero-sum game. Although it is difficult to evaluate the efficiency and practicality of their procedures, the method has an appealing geometric setting.

**§ 2.5.** Completely mixed games of order  $2 \times 2$  were fully investigated by von Neumann [242]. Kaplansky [132] exhibited several special examples of general c.m. games, and extensions of these results were obtained by Bohnenblust, Karlin, and Shapley [36]. Cyclic games constitute a special subclass of games that are invariant with respect to a group of transformations (see also Volume II, Chapter 6 of the present work).

**§ 2.6.** The first method of symmetrizing a game was proposed by Gale, Kuhn, and Tucker [98], the second by Brown and von Neumann [42].

**§ 2.7.** The result of Problem 18 was noted by M. Woodbury, and that of Problem 19 by Sherman [220]. Problem 20 originated with Kaplansky [132].

## CHAPTER 3\*

### DIMENSION RELATIONS FOR SETS OF OPTIMAL STRATEGIES

In this chapter we investigate the structure of sets of solutions of discrete matrix games. Every such set is known to be a convex set, spanned by only a finite number of extreme-point optimal strategies each of which corresponds to a submatrix with special properties (see Theorem 2.4.3). In the following pages a fundamental relationship is established between the dimensions of the sets of optimal strategies. Particular attention is directed to the class of games whose solutions are unique. For example, it is shown that the class of all games of a given order with unique solutions constitutes an open and dense set within the set of all possible matrix games of the same order. We also solve the problem of constructing a game matrix possessing specified solutions.

**3.1 The principal theorems.** Before stating the principal theorems we introduce the following notation. Let the vector  $\mathbf{a}_i$  denote the  $i$ th row of the pay-off matrix  $\|a_{ij}\|$ , and let the vector  $\mathbf{b}_j$  denote the  $j$ th column. As previously,  $\mathbf{e} = \langle 1, 1, \dots, 1 \rangle$ .

Let  $X^0$  and  $Y^0$  denote the sets of optimal strategies for Players I and II, respectively.

Let  $I_1$  be the set of all row indices  $i$  with the property that for some optimal strategy  $\mathbf{x}^{(i)}$ ,

$$x_i^{(i)} > 0.$$

Let  $I_2$  be the set of all row indices  $i$  with the property that for every optimal strategy  $\mathbf{y}$ ,

$$\sum_{j=1}^m a_{ij} y_j = v.$$

In terms of our previous terminology, the set  $I_2$  consists of all rows which are equalizers with respect to all optimal  $\mathbf{y}$  strategies.

Let  $J_1$  and  $J_2$  be the corresponding sets of column indices  $j$  satisfying the same properties with respect to the optimal strategies of Player II. More generally, for any strategy  $\mathbf{x}$  let  $I_1(\mathbf{x})$  be the set of indices  $i$  for which

---

\* This entire chapter may be omitted at first reading without loss of continuity.

$x_i > 0$ , and  $I_2(\mathbf{y})$  the set of indices  $i$  for which  $(\mathbf{a}_i, \mathbf{y}) = v$ , and similarly with  $J_1(\mathbf{y})$  and  $J_2(\mathbf{x})$ . It follows that

$$I_1 = \bigcup_{x \in X^0} I_1(\mathbf{x}), \quad I_2 = \bigcap_{y \in Y^0} I_2(\mathbf{y}),$$

$$J_1 = \bigcup_{y \in Y^0} J_1(\mathbf{y}), \quad J_2 = \bigcap_{x \in X^0} J_2(\mathbf{x}).$$

Let  $\bar{X}$  denote the set of all  $\mathbf{x}$  with  $I_1(\mathbf{x}) \subseteq I_1$  and  $\bar{Y}$  denote the set of all  $\mathbf{y}$  with  $J_1(\mathbf{y}) \subseteq J_1$ . Thus  $\bar{X}$  is the smallest face of the fundamental simplex of mixed strategies containing  $X^0$ , etc. Our immediate objective is to prove the following two relations:

► **THEOREM 3.1.1.**  $I_1 = I_2, J_1 = J_2$ .

► **THEOREM 3.1.2.**  $\dim \bar{X} - \dim X^0 = \dim \bar{Y} - \dim Y^0$ .

Several different types of proofs for Theorems 3.1.1 and 3.1.2 have been given. One proof of Theorem 3.1.2 is based upon representing optimal strategies for Player I as separating hyperplanes and those of Player II as contact points between the two special convex sets, as described in Remark 1.4 of Chapter 1. By carefully counting the number of linearly independent separating planes and noting the dimensions of the contact set, the dimension relations can be established. Another proof of both theorems is based on the proof of the min-max theorem which uses the Kakutani fixed-point theorem; this line of argument will be followed in Chapter 3 of Vol. II, where we consider an extension of these structure theorems to a more general class of games. A third proof, the one given in this chapter, is essentially algebraic in nature, although it has some connections with the geometric proof based on separating convex bodies by hyperplanes.

**3.2 Proof of Theorem 3.1.1.** The integer-valued function  $\omega(\mathbf{x})$  can be used to denote the number of positive components of any strategy  $\mathbf{x}$ . An optimal strategy  $\mathbf{x}$  is said to have full weight if  $x_i > 0$  for every  $i$  in  $I_1$ . Note that if  $\mathbf{x}$  has full weight, then  $\omega(\mathbf{x}) = n(I_1)$ , where  $n(I_1)$  is the number of elements of  $I_1$ .

► **LEMMA 3.2.1.**  $I_1 \subset I_2, J_1 \subset J_2$ .

This lemma follows immediately from Lemma 2.1.2.

► **LEMMA 3.2.2.** There exists an optimal strategy  $\mathbf{x}$  such that (1) for every  $i$  in  $I_1$ ,  $x_i > 0$ ; (2) the vector  $\mathbf{x}\mathbf{A}$  equals  $v$  for all components corresponding to columns in  $J_2$ , and for no others.

*Proof.* By the definition of  $I_1$ , for every  $i \in I_1$  there exists  $\mathbf{x}^{(i)}$  optimal such that  $x_i^{(i)} > 0$ . Let  $\mathbf{x}^0$  be the average of  $\mathbf{x}^{(i)}$ . Then  $x_i^0 > 0$  for every  $i \in I_1$ , and  $\mathbf{x}^0$  is optimal (Lemma 2.1.1).

Furthermore, for every  $j \notin J_2$  there exists  $\mathbf{x}^{(j)}$  optimal such that  $(\mathbf{x}^{(j)} \mathbf{A})_j > v$ . Let  $\mathbf{x}^{00}$  be the average of  $\mathbf{x}^{(j)}$ ;  $\mathbf{x}^{00}$  is optimal and satisfies condition (2). Finally, we set

$$\mathbf{x} = \frac{\mathbf{x}^0 + \mathbf{x}^{00}}{2}.$$

Then

- (1)  $\mathbf{x}$  is optimal since  $X^0$  is convex;
- (2)  $x_i > 0$  for every  $i \in I_1$ ;
- (3)  $(\mathbf{x} \mathbf{A})_j = v$  for  $j \in J_2$ , (by the definition of  $J_2$ );
- (4)  $(\mathbf{x} \mathbf{A})_j > v$  for  $j \notin J_2$ .

*Reduced games.* By the reduced game we mean the game whose pay-off matrix  $\tilde{\mathbf{A}}$  is the submatrix of  $\mathbf{A}$  that keeps only the rows of  $\mathbf{A}$  with indices ranging over  $I_2$  and the columns with indices ranging over  $J_2$ . The reduced game is also called the essential part of the game.

The sets of optimal strategies for the reduced game will be denoted by  $X_0^0, Y_0^0$ .

We observe that  $\tilde{\mathbf{A}}$  and  $\mathbf{A}$  have the same value  $v$ , and that  $X^0 \subset X_0^0, Y^0 \subset Y_0^0$ . To verify these statements, let  $\mathbf{x}, \mathbf{y}$  be a pair of optimal strategies for the game  $\mathbf{A}$ . It follows from the definitions of  $I_1$  and  $J_1$  that  $\mathbf{x}$  and  $\mathbf{y}$  can have positive probability only on those components corresponding to rows and columns of  $\tilde{\mathbf{A}}$ . Thus they can be considered as strategies for the game  $\tilde{\mathbf{A}}$ .

Furthermore, if  $\mathbf{x}, \mathbf{y}$  are optimal for  $\mathbf{A}$ , then  $\mathbf{x} \tilde{\mathbf{A}} = \mathbf{v}$  and  $\tilde{\mathbf{A}} \mathbf{y} = \mathbf{v}$  by the definitions of  $I_2$  and  $J_2$ . Interpreting these two equations with respect to the game  $\tilde{\mathbf{A}}$ , we find that  $\mathbf{x}, \mathbf{y}$  are optimal for  $\tilde{\mathbf{A}}$ , and that  $v(\tilde{\mathbf{A}}) = v(\mathbf{A}) = v$ .

►LEMMA 3.2.3.

$$\dim(X^0) = \dim(X_0^0), \dim(Y^0) = \dim(Y_0^0). \quad (3.2.1)$$

*Proof.* Select  $\mathbf{x}$  in  $X^0$  satisfying the conditions of Lemma 3.2.2. Choose  $\mathbf{x}'$  in  $X_0^0$  but not in  $X^0$ . Then  $\mathbf{x}' \mathbf{A} \geq \mathbf{v}$  for all columns in  $J_2$ ,  $\mathbf{x} \mathbf{A} > \mathbf{v}$  for all columns not in  $J_2$ , and  $\mathbf{x} \mathbf{A} = \mathbf{v}$  for all columns in  $J_2$ . Set  $\bar{\mathbf{x}} = \epsilon \mathbf{x}' + (1 - \epsilon) \mathbf{x}$ . Then  $\bar{\mathbf{x}} \mathbf{A} \geq \mathbf{v}$  for sufficiently small  $\epsilon$ , so that  $\bar{\mathbf{x}}$  is also an optimal strategy of  $X^0$ . It follows from the definition of  $J_2$  that  $\mathbf{x}' \mathbf{A} = \mathbf{v}$  for all columns in  $J_2$ .

Since  $\mathbf{x}'$  is on the extension of the line segment connecting  $\mathbf{x}$  with  $\bar{\mathbf{x}}$ , we can characterize  $X_0^0$  as the part of the linear extension of  $X^0$  which is

contained in the fundamental simplex of strategies  $S^n$  for Player I. Therefore  $\dim(X^0) = \dim(X_0^0)$ . The same argument is valid for  $Y^0$ .

Our aim is to prove that  $I_1 = I_2$  and  $J_1 = J_2$ . Since the sets  $I_1, J_1, I_2, J_2$  are the same for the reduced game  $\tilde{\mathbf{A}}$  as for the original game  $\mathbf{A}$ , we may without loss of generality restrict ourselves to the reduced game. It is clear from the construction of  $\tilde{\mathbf{A}}$  that  $I_2$  and  $J_2$  consist of all  $\tilde{\mathbf{A}}$ 's row indices and column indices, respectively. In order to complete the proof of Theorem 3.1.1 it is sufficient, by virtue of Lemma 3.2.3, to exhibit a strategy  $\mathbf{x}$  in the game  $\tilde{\mathbf{A}}$  such that all components of  $\mathbf{x}$  are positive, and to find the corresponding optimal strategy  $\mathbf{y}$ . To this end, we introduce a family of games depending on a parameter defined by matrices

$$\tilde{\mathbf{A}}(\alpha) \quad (0 \leq \alpha < 1),$$

where the  $i$ th row  $\mathbf{a}_i(\alpha)$  of  $\tilde{\mathbf{A}}(\alpha)$  is given by

$$\begin{aligned} \mathbf{a}_i(\alpha) &= (1 - \alpha)\mathbf{a}_i + \alpha \left( \frac{1}{n} \sum_{i=1}^n \mathbf{a}_i \right) \\ &= (1 - \alpha)\mathbf{a}_i + \alpha \bar{\mathbf{a}}. \end{aligned}$$

Let  $v(\alpha)$  be the value of the game  $\tilde{\mathbf{A}}(\alpha)$ . Then  $v(\alpha) \leq v$ . For let  $\mathbf{y}$  be any strategy in  $Y^0$ . A direct calculation yields

$$\begin{aligned} \tilde{\mathbf{A}}(\alpha)\mathbf{y} &= \left\langle (1 - \alpha) \sum_j a_{ij} y_j + \alpha \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m a_{ij} y_j \right\rangle \\ &\leq (1 - \alpha)\mathbf{v} + \frac{\alpha}{n} n\mathbf{v} = \mathbf{v}. \end{aligned}$$

Hence  $v(\alpha) \leq v$ .

The key auxiliary result needed to complete the proof of Theorem 3.1.1 is embodied in the following lemma.

► **LEMMA 3.2.4.** If there exists an optimal strategy  $\mathbf{y}$  in  $Y_0^0$  which is also in the simplex  $Y[J_1(\alpha)]$  [the convex hull of the pure strategies with indices in  $J_1(\alpha)$ ] for some  $\alpha$ , then  $v(\alpha) = v$ .

*Proof.* Construct  $\mathbf{y}'$  satisfying the conditions of Lemma 3.2.2 with respect to the  $\tilde{\mathbf{A}}(\alpha)$  game. By construction  $\mathbf{y}'$  is an interior point of  $Y[J_1(\alpha)]$ , since it has full weight. Choose  $\epsilon > 0$  sufficiently small so that  $\mathbf{y}'' = (1 + \epsilon)\mathbf{y}' - \epsilon\mathbf{y}$  is still in  $Y[J_1(\alpha)]$ . Note first that

$$v(\alpha) \leq \max_i \{((1 - \alpha)\mathbf{a}_i + \alpha \bar{\mathbf{a}}, (1 + \epsilon)\mathbf{y}' - \epsilon\mathbf{y})\}.$$

Since  $\mathbf{y}$  is optimal and  $i$  is in  $I_2$ , and since the indices in  $I_2$  coincide with

all the rows of  $\tilde{\mathbf{A}}(\alpha)$ , we have  $(\mathbf{a}_i, \mathbf{y}) = v$  and  $(\bar{\mathbf{a}}, \mathbf{y}) = v$  independent of  $i$ . Consequently,

$$\begin{aligned} v(\alpha) &\leq \max_i \{(1 + \epsilon)[(1 - \alpha)(\mathbf{a}_i, \mathbf{y}') + \alpha(\bar{\mathbf{a}}, \mathbf{y}')] - \epsilon v\} \\ &\leq (1 + \epsilon)v(\alpha) - \epsilon v, \end{aligned}$$

since  $\mathbf{y}'$  is optimal for  $\tilde{\mathbf{A}}(\alpha)$ . Then  $\epsilon v \leq \epsilon v(\alpha)$ , or  $v \leq v(\alpha)$ ; and since the reverse inequality always holds,  $v = v(\alpha)$ . The proof of the lemma is complete.

We shall need the fact that if  $\mathbf{y}^\delta \rightarrow \mathbf{y}$ , then  $\tilde{\mathbf{A}}(\alpha_\delta)\mathbf{y}^\delta \rightarrow \tilde{\mathbf{A}}\mathbf{y}$  (where  $\alpha_\delta \rightarrow 0$ ). This statement is obvious and will not be formally proved.

► **LEMMA 3.2.5.** There exists an  $\alpha > 0$  such that  $v(\alpha) = v$ .

*Proof.* We must show that the conditions of Lemma 3.2.4 are satisfied.

Let  $\alpha_\delta$  be a sequence converging to zero. Since  $\tilde{\mathbf{A}}$  has a finite number of columns, there exists an infinite number of  $\alpha_\delta$ 's such that the sets  $J_1(\alpha_\delta)$  agree. We shall restrict ourselves to this subsequence without changing notation.

Select  $\mathbf{y}_\delta$  so that it is optimal for the game  $\tilde{\mathbf{A}}(\alpha_\delta)$ . Then  $\mathbf{y}_\delta$  is in  $Y[J_1(\alpha_\delta)]$ . Since  $Y$  is compact, we may select  $\mathbf{y}$  as a limit point for  $\mathbf{y}_\delta$ . Then  $\mathbf{y}$  is in  $Y[J_1(\alpha_\delta)]$  for all  $\delta$ .

It remains to be shown that  $\mathbf{y}$  is optimal for the game  $\mathbf{A}$ . Clearly,  $\tilde{\mathbf{A}}(\alpha_\delta)\mathbf{y}_\delta \leq \mathbf{v}(\alpha_\delta) \leq \mathbf{v}$ , and therefore  $\tilde{\mathbf{A}}\mathbf{y} \leq \mathbf{v}$ ; hence  $\mathbf{y}$  is optimal and in  $Y_0^0$ . Applying Lemma 3.2.4, we obtain  $v = v(\alpha_\delta)$ .

*Proof of Theorem 3.1.1.* Select  $\alpha$  small enough so that  $v(\alpha) = v$ , and choose  $\mathbf{x}$  optimal for  $\tilde{\mathbf{A}}(\alpha)$ . Form the mixed strategy

$$\mathbf{x}^0 = (1 - \alpha)\mathbf{x} + \frac{\alpha}{n}\mathbf{e},$$

which has full weight (every component is positive). We now show that  $\mathbf{x}^0$  is optimal:

$$\begin{aligned} (\mathbf{x}^0 \tilde{\mathbf{A}})_j &= \sum_{i=1}^n x_i^0 a_{ij} = (1 - \alpha) \sum_{i=1}^n x_i a_{ij} + \frac{\alpha}{n} \sum_{i=1}^n a_{ij} \\ &= (1 - \alpha) \sum_{i=1}^n x_i a_{ij} + \frac{\alpha}{n} \left( \sum_{i=1}^n a_{ij} \right) \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n \left[ (1 - \alpha) a_{ij} + \frac{\alpha}{n} \left( \sum_{k=1}^n a_{kj} \right) \right] x_i \geq v(\alpha) = v. \end{aligned}$$

This implies that  $I_1 = I_2$ . The proof of  $J_1 = J_2$  is similar.

**3.3 Proof of Theorem 3.1.2.** In this section we make use of the mapping properties of a matrix. For the general  $n \times m$  matrix  $\mathbf{B}$ ,  $R(\mathbf{B})$  [or sometimes  $\mathcal{R}(\mathbf{B})$ ] designates the range of the mapping  $\mathbf{B}$  viewed as a linear transformation, and  $\eta(\mathbf{B})$  the linear space of zeros of  $\mathbf{B}$ .

The orthogonal complement of a set  $L$  in Euclidean space is denoted by  $L^\perp$ .

Finally, if  $\mathbf{B}$  is a linear transformation mapping the space  $E^m$  into  $E^n$ , the following orthogonality relations hold:

- (1)  $n = \dim R(\mathbf{B}) + \dim R(\mathbf{B})^\perp$ ;
- (2)  $m = \dim R(\mathbf{B}') + \dim R(\mathbf{B}')^\perp$     ( $\mathbf{B}'$  is the transpose of  $\mathbf{B}$ );
- (3) if  $r(\mathbf{B})$  is the rank of  $\mathbf{B}$ , then  $r(\mathbf{B}') = r(\mathbf{B}) = \dim R(\mathbf{B}) = \dim R(\mathbf{B}')$ .

For a fuller discussion of the matrix algebra, we refer the reader to Sections A.3 and A.5 of Appendix A.

Continuing our discussion of dimension relations for sets of optimal strategies, we now establish an important auxiliary theorem. In all that follows,  $\tilde{\mathbf{A}}$  denotes the matrix of the essential part of the game with matrix  $\mathbf{A}$ .

►**THEOREM 3.3.1.** Let  $X^0$  and  $Y^0$  denote the spaces of optimal strategies for Players I and II, respectively. Then (i) if  $v \neq 0$ ,

$$\dim Y^0 = \dim \eta(\tilde{\mathbf{A}}), \dim X^0 = \dim \eta(\tilde{\mathbf{A}}');$$

and (ii) if  $v = 0$ ,

$$\dim Y^0 + 1 = \dim \eta(\tilde{\mathbf{A}}), \dim X^0 + 1 = \dim \eta(\tilde{\mathbf{A}}').$$

*Proof of (i).* We recall that if  $\mathbf{x} \in X^0$ , the only nonzero components of  $\mathbf{x}$  are those that occur in the rows of  $\tilde{\mathbf{A}}$ . Hence we may consider  $\mathbf{x}$  as a vector in the Euclidean space of dimension equal to the number of rows of  $\tilde{\mathbf{A}}$ . Similarly for  $\mathbf{y} \in Y^0$ .

Let  $\mathbf{y}^0 \in Y^0$  be fixed. Let  $\mathbf{y}^1, \dots, \mathbf{y}^k$  be a maximal set of vectors in  $Y^0$  such that  $\mathbf{y}^1 - \mathbf{y}^0, \dots, \mathbf{y}^k - \mathbf{y}^0$  are linearly independent. It follows from the definition of the essential part of the game that  $\tilde{\mathbf{A}}\mathbf{y}^0 \equiv \mathbf{v}$  and  $\tilde{\mathbf{A}}\mathbf{y}^i \equiv \mathbf{v}$ ; hence  $\tilde{\mathbf{A}}(\mathbf{y}^i - \mathbf{y}^0) \equiv 0$ , and  $\{\mathbf{y}^i - \mathbf{y}^0\}$  constitutes a set of linearly independent zeros of  $\tilde{\mathbf{A}}$ . Therefore  $\dim Y^0 \leq \dim \eta(\tilde{\mathbf{A}})$ .

On the other hand, by Lemma 3.2.2 and Theorem 3.1.1, there exists an optimal strategy  $\mathbf{y}^0$  such that  $y_j^0 > 0$  for all  $j$  ( $j$  ranging over the columns of  $\tilde{\mathbf{A}}$ ). Let  $\tilde{\mathbf{A}}\mathbf{z} = 0$ . We argue as in the proof of Lemma 2.4.1: Let  $\mathbf{x}^0$  be optimal such that  $\mathbf{x}^0\tilde{\mathbf{A}} = \mathbf{v}$ ; then  $0 = (\mathbf{x}^0, \tilde{\mathbf{A}}\mathbf{z}) = (\mathbf{x}^0\tilde{\mathbf{A}}, \mathbf{z}) = v\sum z_i$ , and hence  $\sum z_i = 0$ . Consequently, for  $\epsilon$  sufficiently small  $\mathbf{y}^0 + \epsilon\mathbf{z}$  is an optimal strategy. Hence  $\dim \eta(\tilde{\mathbf{A}}) \leq \dim Y^0$ . The proof of the second relation in (i) is similar.

*Proof of (ii).* If  $v = 0$ , then a strategy  $\mathbf{y}$  is optimal if and only if  $\tilde{\mathbf{A}}\mathbf{y} = \mathbf{0}$ . That is, any zero of  $\tilde{\mathbf{A}}$  which is also a strategy is an optimal strategy. Therefore  $Y^0$  can be obtained from the set  $\eta(\tilde{\mathbf{A}})$  by imposing the extra relation  $\sum y_i = 1$  intersected with the first orthant. This requirement lowers the dimension by 1. Hence  $\dim Y^0 = \dim \eta(\tilde{\mathbf{A}}) - 1$ . Similar proofs hold for  $X^0$ . Thus the lemma is established.

We are now ready to prove Theorem 3.1.2, which states that

$$\dim \bar{X} - \dim X^0 = \dim \bar{Y} - \dim Y^0.$$

(See page 64 for definitions of  $\bar{X}$  and  $\bar{Y}$ .)

*Proof of Theorem 3.1.2.* By Lemma 3.2.3,  $\dim X^0 = \dim X_0^0$ , where  $X_0^0$  denotes the space of optimal strategies for the reduced game; therefore we may restrict ourselves to the reduced game with matrix  $\tilde{\mathbf{A}}$ .

Since  $\mathbf{y} \in \eta(\tilde{\mathbf{A}}) \leftrightarrow \tilde{\mathbf{A}}\mathbf{y} = \mathbf{0} \leftrightarrow (\mathbf{x}, \tilde{\mathbf{A}}\mathbf{y}) = 0 \leftrightarrow (\mathbf{x}\tilde{\mathbf{A}}, \mathbf{y}) = 0$  for all  $\mathbf{x}$ , we must have

$$\eta(\tilde{\mathbf{A}}) = [R(\tilde{\mathbf{A}}')]^\perp; \quad (3.3.1)$$

similarly,

$$\eta(\tilde{\mathbf{A}}') = [R(\tilde{\mathbf{A}})]^\perp. \quad (3.3.2)$$

Furthermore, since  $\tilde{\mathbf{A}}$  can be viewed as a mapping of the smallest linear space  $\mathcal{K}$  containing  $\bar{Y}$  into the corresponding linear space  $\mathcal{G}$  which contains  $\bar{X}$ , and since  $\tilde{\mathbf{A}}'$  is the corresponding dual transformation of  $\mathcal{G}$  into  $\mathcal{K}$ , it follows from the orthogonality dimension relations (1) and (2) that  $\dim \mathcal{G} = \dim \bar{X} + 1 = \dim R(\tilde{\mathbf{A}}) + \dim [R(\tilde{\mathbf{A}})]^\perp$  and  $\dim \mathcal{K} = \dim \bar{Y} + 1 = \dim R(\tilde{\mathbf{A}}') + \dim [R(\tilde{\mathbf{A}}')]^\perp$ . Using property (3), equations (3.3.1) and (3.3.2), and Theorem 3.3.1, we get the successive relations

$$\begin{aligned} \dim \bar{Y} &= r(\tilde{\mathbf{A}}') + \dim \eta(\tilde{\mathbf{A}}) - 1, \\ \dim \bar{X} &= r(\tilde{\mathbf{A}}) + \dim \eta(\tilde{\mathbf{A}}') - 1, \\ \dim \bar{Y} - \dim \eta(\tilde{\mathbf{A}}) &= \dim \bar{X} - \dim \eta(\tilde{\mathbf{A}}'), \\ \dim \bar{Y} - \dim Y^0 &= \dim \bar{X} - \dim X^0. \end{aligned} \quad (3.3.3)$$

**3.4 The converse of Theorem 3.1.2.** If the optimal strategies for both players are unique, say  $\mathbf{x} = \langle x_1, \dots, x_s \rangle$  and  $\mathbf{y} = \langle y_1, \dots, y_r \rangle$ , then  $\dim X^0 = \dim Y^0 = 0$ . By Theorem 3.1.2,  $\dim \bar{X} = \dim \bar{Y}$ , and hence  $r = s$ . In this case, both strategies must have the same number of positive components.

Let us consider the converse problem of prescribing single strategies  $\mathbf{x}^0$  for Player I and  $\mathbf{y}^0$  for Player II and seeking to construct a matrix  $\mathbf{A}$  for

which  $\mathbf{x}^0$  and  $\mathbf{y}^0$  are the unique optimal strategies. As indicated, we must require that  $\mathbf{x}^0$  and  $\mathbf{y}^0$  have the same number of nonzero components.

Letting  $\mathbf{x}^0 = \langle x_0, x_1, \dots, x_k \rangle$  and  $\mathbf{y}^0 = \langle y_0, y_1, \dots, y_k \rangle$  with all  $x_i$  and  $y_j$  positive, we shall exhibit a matrix of order  $(k+1) \times (k+1)$  with  $\mathbf{x}^0$  and  $\mathbf{y}^0$  the unique optimal strategies. Explicitly, take  $v \neq 1/k$  and let

$$\mathbf{A} = \left( \begin{array}{cccccc} \frac{v(x_0 + y_0 - 1) + \sum_{i=1}^k x_i y_i}{x_0 y_0}, & \frac{v - x_1}{x_0}, & \frac{v - x_2}{x_0}, & \dots, & \frac{v - x_k}{x_0} \\ \frac{v - y_1}{y_0}, & 1, & 0, & \dots, & 0 \\ \frac{v - y_2}{y_0}, & 0, & 1, & \dots, & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{v - y_k}{y_0}, & 0, & 0, & \dots, & 1 \end{array} \right)$$

It is readily verified that  $\mathbf{x}^0$  and  $\mathbf{y}^0$  are optimal strategies, and that the value of the game is  $v$ . It remains to establish the uniqueness.

Let  $\mathbf{w} = \langle w_0, w_1, \dots, w_n \rangle$  be an optimal strategy for Player I. If  $j \neq 0$ , then by Lemma 2.1.2

$$\frac{v - x_j}{x_0} w_0 + w_j = v,$$

and therefore

$$\sum_{j=1}^k \frac{v - x_j}{x_0} w_0 + \sum_{j=1}^k w_j = kv.$$

But

$$\sum_{j=1}^k x_j = 1 - x_0, \quad \sum_{j=1}^k w_j = 1 - w_0,$$

so that

$$(vk - 1 + x_0)w_0 + x_0(1 - w_0) = kvx_0$$

or, equivalently,

$$(vk - 1)w_0 = (vk - 1)x_0.$$

Since  $vk \neq 1$ , it follows that  $w_0 = x_0$ . Further, since

$$\frac{v - x_j}{x_0} w_0 + w_j = v,$$

$w_j = x_j$  for all  $j$ . Thus  $\mathbf{x}^0$  is the only optimal strategy for Player I. A similar argument shows that  $\mathbf{y}^0$  is the only optimal strategy for Player II.

We now set about extending this result. The more general problem is this: Which strategy sets  $X^0$  and  $Y^0$  can be prescribed in advance such

that a matrix  $\mathbf{A}$  can be found for which  $X^0$  and  $Y^0$  are the precise sets of optimal strategies for Players I and II, respectively? A limitation on the choice of  $X^0$  and  $Y^0$  is implied in the dimension relations stated by Theorem 3.1.2. Remarkably enough, however, this is the only restriction. In other words, any sets  $X^0$  and  $Y^0$  satisfying Theorem 3.1.2 for appropriate  $\bar{X}$  and  $\bar{Y}$  can qualify as the sets of optimal strategies for a matrix  $\mathbf{A}$ . A more precise statement of this result follows.

► **THEOREM 3.4.1.** Let  $\bar{X}$  and  $\bar{Y}$  be two fundamental simplexes in  $E^s$  and  $E^r$ , respectively, described by the relations

$$\sum_{i=1}^s x_i = 1, \quad x_i \geq 0, \quad \text{and} \quad \sum_{j=1}^r y_j = 1, \quad y_j \geq 0.$$

Let  $X^0$  and  $Y^0$  be two closed, convex, polyhedral sets contained in  $\bar{X}$  and  $\bar{Y}$ , respectively, and intersecting the interior. If

$$\dim \bar{X} - \dim X^0 = \dim \bar{Y} - \dim Y^0 = q, \quad (3.4.1)$$

then there exists a matrix  $\mathbf{A}$  for which the sets of optimal strategies are  $X^0$  and  $Y^0$  and the value of the game is zero.

*Proof.* Let  $X_0^0$  and  $Y_0^0$  be the linear extensions of  $X^0$  and  $Y^0$  inside  $\bar{X}$  and  $\bar{Y}$ . We shall first construct a matrix  $\tilde{\mathbf{A}}$  with optimal strategies consisting of the sets  $X_0^0$  and  $Y_0^0$ .

To this end, let  $\mathcal{L}_0$  and  $\mathcal{K}_0$  be the smallest linear subspaces containing  $X_0^0$  and  $Y_0^0$ , respectively. Since zero belongs to any linear space, we have  $\dim \mathcal{L}_0 = \dim X_0^0 + 1 = \dim X^0 + 1$  and  $\dim \mathcal{K}_0 = \dim Y_0^0 + 1 = \dim Y^0 + 1$ . It follows from (3.4.1) and orthogonality relations (1) and (2) of Section 3.3 that  $\dim \mathcal{L}_0^\perp = \dim \mathcal{K}_0^\perp = q$ .

Let  $\mathbf{P}$  be the  $s \times s$  matrix that represents a projection of  $E^s$  onto  $\mathcal{L}_0^\perp$ . Let  $\mathbf{Q}$  be a matrix of rank  $q$  which maps  $E^r$  into  $E^s$  and which is a one-to-one linear transformation of  $\mathcal{K}_0^\perp$  onto  $\mathcal{L}_0^\perp$  satisfying  $\eta(\mathbf{Q}) = \mathcal{K}_0$ . Such a matrix exists by virtue of the fact that  $\mathcal{K}_0^\perp$  and  $\mathcal{L}_0^\perp$  are of the same dimension. Set  $\tilde{\mathbf{A}} = \mathbf{PQ}$ . The matrix  $\tilde{\mathbf{A}}$  is an  $s \times r$  matrix.

The zero manifold of  $\tilde{\mathbf{A}}$  is  $\mathcal{K}_0$ , and that of  $\tilde{\mathbf{A}}'$  is  $\mathcal{L}_0$ . These assertions are established as follows: Since  $\mathbf{Q}(\mathcal{K}_0) = 0$  (this means the image of  $\mathcal{K}_0$  by the linear mapping  $\mathbf{Q}$ ), we deduce that  $\eta(\tilde{\mathbf{A}}) \supset \mathcal{K}_0$ . On the other hand, since  $\mathbf{Q}$  is a one-to-one linear transformation of  $\mathcal{K}_0^\perp$  onto  $\mathcal{L}_0^\perp$  and  $\mathbf{P}$  acts as the identity mapping on  $\mathcal{L}_0^\perp$ , we find that  $\tilde{\mathbf{A}}\mathbf{y} = \mathbf{0}$  for  $\mathbf{y}$  in  $\mathcal{K}_0^\perp$  implies that  $\mathbf{y} = \mathbf{0}$ . Since  $\mathcal{K}_0$  and  $\mathcal{K}_0^\perp$  span  $E^r$  as a direct sum, we infer that  $\eta(\tilde{\mathbf{A}}) = \mathcal{K}_0$ . It follows that  $\eta(\tilde{\mathbf{A}}') = R(\tilde{\mathbf{A}})^\perp = (\mathcal{L}_0^\perp)^\perp = \mathcal{L}_0$ , and thus our assertions are valid.

Since the value of the game is zero, the optimal strategies for  $\tilde{\mathbf{A}}$  are precisely the vectors in  $X_0^0$  and  $Y_0^0$ .

We shall now enlarge  $\tilde{\mathbf{A}}$  in order to eliminate those strategies of  $X_0^0$  and  $Y_0^0$  which are not in  $X^0$  and  $Y^0$ . The following characterization of convex polyhedral sets is needed.

Let  $F$  be a closed, convex, polyhedral set of  $E^n$  which is also contained in the simplex

$$x_i \geq 0, \quad \sum_{i=1}^n x_i = 1.$$

The dimension of  $F$  is at most  $n - 1$ , and the dimension of any face of  $F$  is at most  $n - 2$ . Let  $k$  denote the number of faces of  $F$ . By assumption,  $k$  is finite.

For any given face of  $F$ , there exists at least one  $(n - 1)$ -dimensional hyperplane  $B$  of  $E^n$  through the origin which is a supporting plane for  $F$  and contains the given face of  $F$ . Since the dimension of any face of  $F$  is at most  $n - 2$ , such a hyperplane can clearly be found. Let the direction numbers of the normal to the plane be denoted by  $\langle b_1, \dots, b_n \rangle$ . The plane  $B$  separates  $E^n$  into two half-spaces  $E_0$  and  $E_1$ , where  $E_0$  is closed and contains the given face of  $F$ , and  $E_1 = E_0^c$  (the complement of  $E_0$ ). Then

$$\mathbf{x} \in B \rightarrow \sum_{i=1}^n x_i b_i = 0,$$

$$\mathbf{x} \in E_0 \rightarrow \sum_{i=1}^n x_i b_i \geq 0,$$

$$\mathbf{x} \in E_1 \rightarrow \sum_{i=1}^n x_i b_i < 0.$$

Let  $B^1, \dots, B^l$  be a set of half-spaces determined by a collection of supporting hyperplanes for the faces of  $F$  whose intersection is precisely  $F$ . Formally, if  $B^j \rightarrow \langle b_1^j, \dots, b_n^j \rangle$ , then

$$\mathbf{z} \in F \rightarrow \sum_{i=1}^n z_i b_i^j \geq 0 = v \quad \text{for every } j \ (1 \leq j \leq l); \quad (3.4.2)$$

while if  $\mathbf{z} \in F^c$ , then there exists a value  $j_0$  ( $1 \leq j_0 \leq l$ ) such that

$$\sum_{i=1}^n z_i b_i^{j_0} < 0 = v. \quad (3.4.3)$$

Thus the planes  $(B^j)$ , together with the conditions (3.4.2) and (3.4.3), completely determine the set  $F$ . We remark that although a change in orientation (taking  $\{-b_i^j\}$  instead of  $\{b_i^j\}$  for all  $i, j$ ) reverses all the inequalities in (3.4.2) and (3.4.3), the set  $F$  remains completely characterized.

*Construction of  $\mathbf{A}$ .* First we construct the set  $\{B^j\}$  of planes corresponding to the faces of  $X^0$ ; we add  $l$  new columns to  $\tilde{\mathbf{A}}$ , the  $j$ th column added consisting of the vector  $\langle b_1^j, \dots, b_s^j \rangle$ . Next we form the set  $\{C^i\}$  of planes corresponding to the faces of  $Y^0$ , reversing the inequalities in (3.4.2) and (3.4.3); we add  $m$  new rows to  $\tilde{\mathbf{A}}$ , the  $i$ th row consisting of the vector  $\langle c_1^i, \dots, c_r^i \rangle$ . Finally, we complete the matrix  $\mathbf{A}$  by filling in zeros for the lower right corner. If  $\tilde{\mathbf{A}} = \|a_{ij}\|$ , then

$$\mathbf{A} = \begin{vmatrix} a_{11} & \dots & a_{1r} & b_1^1 & \dots & b_1^l \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{s1} & & a_{sr} & b_s^1 & & b_s^l \\ c_1^1 & \dots & c_r^1 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ c_1^m & \dots & c_r^m & 0 & \dots & 0 \end{vmatrix}$$

We shall now verify that  $X^0, Y^0$  are exactly the sets of optimal strategies for  $\mathbf{A}$ . First we note that  $\mathbf{x} \in X^0$  and  $\mathbf{y} \in Y^0$  implies that

$$\mathbf{x} = \langle x_1, \dots, x_s, 0, \dots, 0 \rangle$$

and

$$\mathbf{y} = \langle y_1, \dots, y_r, 0, \dots, 0 \rangle,$$

where

$$\mathbf{x} \in X^0 \rightarrow \sum_{i=1}^s x_i a_{ij} = 0$$

by the construction of  $\tilde{\mathbf{A}}$ , and

$$\mathbf{x} \in X^0 \rightarrow \sum_{i=1}^s x_i b_i^j \geq 0 \quad \text{for all } j \quad (1 \leq j \leq l)$$

by (3.4.2). Hence  $\mathbf{x}$  is optimal for  $\mathbf{A}$ . On the other hand, if  $\mathbf{x} \notin X^0$ , then by (3.4.3) there exists a  $j_0$  such that  $\sum x_i b_i^{j_0} < 0$ ; hence  $\mathbf{x}$  is not optimal. Thus the optimal strategies are exactly the elements of  $X^0$ .

The corresponding argument is valid for  $Y^0$  if we remember that  $\{C^i\}$  was formed with the inequalities reversed in (3.4.2) and (3.4.3). This completes the proof of Theorem 3.4.1.

The matrix constructed in Theorem 3.4.1 possessing prescribed sets of optimal strategies  $X^0$  and  $Y^0$  is not necessarily the most efficient (i.e., the smallest) such matrix that can be constructed. To obtain the most efficient matrix, we must exercise care in the construction of the additional rows and columns. Specifically, there is no need to construct any plane  $B^i$  that

duplicates a face of  $\bar{X}$ . These faces are described by the basic inequalities  $x_i \geq 0$  which define the fundamental simplex. By eliminating these extraneous column and row vectors from the matrix  $\mathbf{A}$  we obtain the smallest matrix with the prescribed sets of optimal strategies. The proof of this assertion is left as an exercise for the reader.

Problems 5 and 6 of this chapter also involve the construction procedure of Theorem 3.4.1.

**3.5 Uniqueness of optimal strategies.** If Players I and II possess unique optimal strategies, the choice of an optimal strategy presents no problem for either player; there is only one to choose. When the sets of optimal strategies are more than zero-dimensional, on the other hand, selecting an optimal strategy involves difficult conceptual questions, the use of additional criteria, and other complications. It is therefore useful and important to know when a game has unique solutions. We shall show in this section that games with unique solutions occur very frequently.

We observe first that for games with unique optimal strategies,  $\dim \bar{X} = \dim \bar{Y}$ . This is an immediate corollary of Theorem 3.1.2. Equivalently, the maximal weight (number of positive components) of the optimal  $\mathbf{x}$  strategy and the maximal weight of the optimal  $\mathbf{y}$  strategy are the same. Hence the reduced matrix is necessarily square.

Next we introduce the following topological structure on the space of all games. By a neighborhood  $U(\mathbf{A})$  of the matrix  $\mathbf{A}$  of order  $n \times m$  we mean the set of all possible matrices of the same order whose components are within a prescribed positive distance of the corresponding components of  $\mathbf{A}$ .

Our principal theorem follows.

► **THEOREM 3.5.1.** If a matrix  $\mathbf{A}$  has unique optimal strategies  $\mathbf{x}$  and  $\mathbf{y}$ , there exists a neighborhood  $U(\mathbf{A})$  such that if  $\mathbf{B}$  is in  $U(\mathbf{A})$ , then  $\mathbf{B}$  has unique optimal strategies.

The proof depends on the following elementary convergence lemma.

► **LEMMA 3.5.1.** Let  $\mathbf{A}_n \rightarrow \mathbf{A}$ . Let  $X^0$  and  $Y^0$  be the sets of optimal strategies for  $\mathbf{A}$ , and let  $X_n^0$  and  $Y_n^0$  be the corresponding sets for the game  $\mathbf{A}_n$ . If  $G$  is an open set containing  $X^0$ , then for  $n$  sufficiently large,  $X_n^0 \subset G$ . Moreover,  $v_n \rightarrow v$ .

*Proof.* The proof of the last statement of the lemma is trivial. Suppose the first part of the lemma is false. Then there exists a subsequence  $\{n_i\}$  and a corresponding sequence  $\{\mathbf{x}^{n_i}\}$  such that  $\mathbf{x}^{n_i}$  is optimal for  $\mathbf{A}_{n_i}$  and is in  $G^c$ , where  $G^c$  is the relative complement of  $G$  within  $S$  (the simplex of all strategies for Player I).

Since  $G^c$  is a closed bounded set, the sequence  $\{\mathbf{x}^{n_i}\}$  must have a subsequence converging to a limit  $\mathbf{x}^0$  in  $G^c$ . Thus, since  $\lim v_n = v$ , we get  $\mathbf{x}^{n_i} \mathbf{A}_{n_i} \geq \mathbf{v}_{n_i}$ , and in the limit,  $\mathbf{x}^0 \mathbf{A} \geq \mathbf{v}$ ; hence  $\mathbf{x}^0$  is in  $G$ . But  $\mathbf{x}^0$  is also in  $G^c$ , which is impossible.

*Proof of Theorem 3.5.1.* Without loss of generality we may assume  $v \neq 0$  (translate all the elements of  $\mathbf{A}$  by the same constant, if necessary). Let  $\mathbf{A}^n \rightarrow \mathbf{A}$ , so that  $v_n \neq 0$  for  $n$  sufficiently large. We define the sets  $I_1, I_2, I_1^n, I_2^n$  as in Section 3.1. Explicitly,

$$I_2^n = \{i \mid \sum_j a_{ij}^n y_j^n = v \quad \text{for all } \mathbf{y}^n \text{ optimal for } \mathbf{A}^n\}.$$

Since the total number of row indices  $i$  of  $\mathbf{A}$  is finite, there can be only a finite number of distinct sets  $I_2^n$ . Let  $I_2^n$  be any set that occurs infinitely often in  $\{I_2^n\}$ .

For every  $i \in I_2^n$  we find a subsequence  $\{n_i\}$  such that

$$\sum_j a_{ij}^{n_i} y_j^{n_i} = v^{n_i},$$

and in the limit,  $\sum_j a_{ij} y_j = v$ , where  $\mathbf{y}$  is the unique optimal strategy for  $\mathbf{A}$ . In fact, by Lemma 3.5.1 every limit point of  $\mathbf{y}^n$  is the same unique optimal strategy  $\mathbf{y}$ . Since  $\mathbf{y}$  is unique optimal for  $\mathbf{A}$ , the last equation shows that  $i \in I_2$ .

Consequently,  $I_2^n \subset I_2$  for all but a finite number of  $n$ .

On the other hand, if  $x_i > 0$  and  $\mathbf{x}^n \rightarrow \mathbf{x}$ , then for  $n$  sufficiently large,  $x_i^n > 0$ , since  $\mathbf{x}$  is the unique optimal strategy for Player I and is thus the only limit point  $X_n^0$  can have (Lemma 3.5.1). Hence  $I_1^n \supset I_1$  for  $n$  large. Combining, we have  $I_2 \supset I_2^n \supset I_1^n \supset I_1$  for  $n$  large. But  $I_2 = I_1$  by Theorem 3.1.1, and therefore  $I_2^n = I_2$  for  $n$  large. A similar argument enables us to show that for  $n$  sufficiently large,  $J_2^n = J_2$ .

As observed above, the reduced game  $\tilde{\mathbf{A}}$  is a square matrix. Furthermore, since  $J_2^n = J_2$  and  $I_2^n = I_2$ , it follows that the reduced games  $\tilde{\mathbf{A}}^n$  are also square matrices, and of the same order as  $\tilde{\mathbf{A}}$ . But by the hypothesis  $\dim X^0 = 0$  and the identity  $1 + \dim \bar{X} - \dim X^0 = \text{rank}(\tilde{\mathbf{A}})$  [see (3.3.3)],  $\tilde{\mathbf{A}}$  has full rank; and since a small perturbation applied to the components can at worst increase the rank of a matrix, it follows that  $\tilde{\mathbf{A}}^n$  must also have full rank for  $n$  sufficiently large. Hence,  $1 + \dim \bar{X}^n = \text{rank}(\tilde{\mathbf{A}}^n)$ . However,  $1 + \dim \bar{X}^n - \dim X_n^0 = \text{rank}(\tilde{\mathbf{A}}^n)$ . Therefore, the only conclusion is that  $\dim X_n^0 = 0$ , that is, that the optimal strategy  $\mathbf{x}^n$  for  $\mathbf{A}^n$  is unique. An analogous argument leads to the same conclusion for the  $\mathbf{y}$  optimal strategies, and the proof is complete.

Theorem 3.5.1 states that the set of all matrices with unique solutions is an open set. The next theorem asserts that the complement of this open

set is sparse, in the sense that any game with nonunique solutions can be approximated elementwise as closely as desired by games possessing unique solutions. It is gratifying to know that most games have unique solutions; it should be pointed out, however, that in practice we seem constantly to be dealing with games having nonunique solutions.

► **THEOREM 3.5.2.** The set of all matrix games with unique solutions constitutes a dense set in the set of all matrix games.

The proof will be obtained with the aid of the following two lemmas, each possessing independent interest.

► **LEMMA 3.5.2.** Suppose that  $\mathbf{B}$  is a matrix with the following properties for all  $r$ : (1) the matrix obtained by adjoining a column of 1's to any  $(r+1) \times r$  submatrix of  $\mathbf{B}$  is nonsingular; (2) the same condition holds for the submatrices of  $\mathbf{B}'$ ; (3) every  $r \times r$  submatrix of  $\mathbf{B}$  is nonsingular. Then the optimal strategies for  $\mathbf{B}$  are unique.

*Proof.* Let  $\mathbf{x}^0$  and  $\mathbf{y}^0$  be optimal strategies. Let  $r$  be the number of  $i$ 's such that  $x_i^0 > 0$ ,  $t$  the number of  $j$ 's such that  $y_j^0 > 0$ ,  $s$  the number of  $j$ 's such that  $\sum_i a_{ij} x_i^0 = v$ , and  $w$  the number of  $i$ 's such that  $\sum_j a_{ij} y_j^0 = v$ . By Lemma 2.1.2,

$$w \geq r \quad \text{and} \quad s \geq t. \quad (3.5.1)$$

Let  $\bar{\mathbf{B}}$  be the  $r \times s$  submatrix of  $\mathbf{B}$  whose rows correspond to the indices  $i$  for which  $x_i^0 > 0$ , and whose columns correspond to the indices  $j$  for which  $\sum_i a_{ij} x_i^0 = v$ .

Let  $\mathbf{C}$  be the matrix of order  $s \times (r+1)$  obtained by adjoining a column of 1's to  $\bar{\mathbf{B}}'$ . Set  $\mathbf{x}^0 = \langle x_1^0, \dots, x_r^0, -v \rangle$ .

Then  $\mathbf{C}\mathbf{x}^0 = \mathbf{0}$ . Suppose that  $r+1 \leq s$ . Then we can select any  $(r+1) \times (r+1)$  submatrix  $\mathbf{D}$  of  $\mathbf{C}$ . For this  $\mathbf{D}$ , we have  $\mathbf{D}\mathbf{x}^0 \equiv \mathbf{0}$ , which is impossible, since  $\mathbf{D}$  is nonsingular by assumption (3).

Therefore

$$r+1 > s. \quad (3.5.2)$$

A similar argument, using the  $w \times t$  matrix associated with the vector  $\mathbf{y}$ , shows that

$$t+1 > w. \quad (3.5.3)$$

Combining (3.5.1) and (3.5.3) gives us  $t+1 > w \geq r$ ; hence  $t \geq r$ . From  $s \geq t$  and  $t \geq r$  we obtain  $s \geq r$ , and finally  $r+1 > s$  gives  $r \geq s$ . It follows readily that  $r = s = t = w$ .

Since  $\mathbf{x}^0$  and  $\mathbf{y}^0$  were arbitrary optimal strategies and satisfy the identity  $t = s$ , we conclude that all optimal  $\mathbf{y}$  strategies have exactly the same number of positive components. Therefore the indices of positive components for two  $\mathbf{y}$  solutions coincide, for otherwise a convex combination

of  $\mathbf{y}$  and  $\mathbf{y}^0$  would lead to an optimal strategy whose weight exceeds  $t$ . Similarly, all optimal  $\mathbf{x}$  strategies possess the same  $t = s$  positive components.

The equation  $r = s$  then indicates that for any two optimal strategies  $\mathbf{x}^1$  and  $\mathbf{x}^2$ ,

$$(\mathbf{x}^1 - \mathbf{x}^2)\bar{\mathbf{B}} = \mathbf{0}.$$

Since  $\bar{\mathbf{B}}$  is nonsingular by assumption,  $\mathbf{x}^1 = \mathbf{x}^2$  and  $\dim X^0 = 0$ . Similarly,  $\dim Y^0 = 0$ .

At this point, we exhibit a matrix satisfying the conditions of Lemma 3.5.2. Choose  $\alpha_1, \dots, \alpha_n$  all positive and distinct ( $\alpha_i \neq 1$ ), and take  $k_1, \dots, k_n$  as distinct positive integers. Consider

$$\mathbf{B} = \begin{vmatrix} \alpha_1^{k_1} & \cdots & \alpha_n^{k_1} \\ \alpha_1^{k_2} & \cdots & \alpha_n^{k_2} \\ \vdots & & \vdots \\ \alpha_1^{k_n} & \cdots & \alpha_n^{k_n} \end{vmatrix} \quad (3.5.4)$$

We now verify that  $\mathbf{B}$  satisfies the conditions of the lemma.

Observe that square submatrices of  $\mathbf{B}$  are of the same form as  $\mathbf{B}$ . If we take an  $r \times (r + 1)$  submatrix of  $\mathbf{B}$  and adjoin a row of 1's, we again get a matrix of the same form, provided that we assign  $k_{n+1} = 0$ . Adjoining a column of 1's to an  $(r + 1) \times r$  submatrix also provides us with a matrix of the same type where we assign  $\alpha_{n+1} = 1$ .

Therefore, it suffices to verify that  $\det \mathbf{B} \neq 0$  for any choice of distinct  $k_i \neq 0$  and  $\alpha_i$ . This is the classical Vandermonde type of determinant, which is always nonvanishing.

► **LEMMA 3.5.3.** The set of all matrices satisfying the conditions of Lemma 3.5.2 is dense in the set of all matrices.

*Proof.* We take  $\mathbf{A}$  arbitrary,  $\mathbf{B}$  of the form exhibited in (3.5.4). Let  $\mathbf{R}$  be any  $r \times r$  submatrix of  $\mathbf{A} + \epsilon \mathbf{B}$ .

The determinant of  $\mathbf{R}$  is a polynomial of exact degree  $r$  in  $\epsilon$ . In fact, the coefficient of  $\epsilon^r$  is the determinant of the corresponding submatrix of  $\mathbf{B}$ , which is not zero. This polynomial has at most  $r$  strictly positive roots.

Similarly, the determinant of every square matrix obtained by adding a row or a column of 1's to an  $r \times (r + 1)$  or an  $(r + 1) \times r$  submatrix of  $\mathbf{A} + \epsilon \mathbf{B}$  is a polynomial of exact degree  $r + 1$  in  $\epsilon$ , and the number of real roots of each such polynomial is at most  $r + 1$ .

The number of all positive roots  $\epsilon_i$  obtained from all possible subdeterminants of  $\mathbf{A} + \epsilon \mathbf{B}$  is finite.

Hence  $\epsilon_0 = \min_i \{\epsilon_i\}$  is strictly positive. If  $0 < \epsilon < \epsilon_0$ , then  $\mathbf{A} + \epsilon \mathbf{B}$  is of type **B** and therefore possesses unique optimal strategies. This completes the proof of the lemma, and the proof of Theorem 3.5.2.

In concluding, it is worth pointing out that Theorem 3.5.1 does not apply to games in which only one player has a unique optimal strategy. For example, let

$$\mathbf{A} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix};$$

then  $Y^0$  is the set  $\{\langle y, 1 - y \rangle\}$  ( $0 \leq y \leq 1$ ), while  $X^0$  is a single strategy  $\mathbf{x}^0 = \langle 1, 0 \rangle$ . Now consider the perturbed game

$$\mathbf{A}_\epsilon = \begin{vmatrix} 1 + \epsilon & 1 \\ 0 & 1 \end{vmatrix},$$

with  $\epsilon > 0$ . In this game  $Y^0(\epsilon)$  is a single strategy  $\mathbf{y}^0 = \langle 0, 1 \rangle$ , while  $X^0(\epsilon)$  consists of all strategies  $\langle x, 1 - x \rangle$ , where  $1/(1 + \epsilon) \leq x \leq 1$ . Hence a small perturbation changes the dimension of  $X^0$  from zero to one.

### 3.6 Problems.

1. Let  $\mathbf{A}$  be a  $3 \times 3$  matrix game all of whose optimal strategies use the same components with positive probability. Show that the solution to this game is unique if the number of components used by the two players is the same.
2. Show that the reduced game of a symmetric game is also symmetric.
3. Show that for symmetric games  $\dim \bar{X} - \dim X^0$  is even.
4. Show that if  $X^0 = Y^0$  and if  $\dim \bar{X} - \dim X^0$  is even, there exists a symmetric game with optimal strategies spanned by  $X^0, Y^0$ .
5. Construct a matrix game with extreme-point solutions

$$\mathbf{x} = \langle 0, 0, 1 \rangle, \quad \mathbf{y}_1 = \langle \frac{1}{3}, \frac{2}{3}, 0 \rangle, \quad \mathbf{y}_2 = \langle \frac{2}{3}, \frac{1}{3}, 0 \rangle.$$

6. Construct a matrix game with extreme-point solutions

$$\begin{aligned} \mathbf{x}_1 &= \langle \frac{1}{2}, \frac{1}{2}, 0 \rangle, & \mathbf{y}_1 &= \langle \frac{1}{3}, 0, 0, \frac{2}{3} \rangle, \\ \mathbf{x}_2 &= \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle; & \mathbf{y}_2 &= \langle 0, \frac{1}{2}, 0, \frac{1}{2} \rangle, \\ & & \mathbf{y}_3 &= \langle 0, 0, \frac{1}{3}, \frac{2}{3} \rangle. \end{aligned}$$

7. Show that the infinite matrix game defined by  $a_{ij} = i - j$  ( $i, j = 1, 2, 3, \dots$ ) has no value.
8. Show that the infinite matrix game defined by

$$a_{ij} = \frac{i - j}{\sqrt{1 + (i - j)^2}} \quad (i, j = 1, 2, \dots)$$

has no value.

9. Show that if  $c_k > 0$  and  $\sum_{k=1}^{\infty} c_k = 1$ , the infinite matrix game

$$\left\| \begin{array}{cccccc} 0 & c_1 - 1 & c_2 & c_3 & \dots \\ 0 & c_1 & c_2 - 1 & c_3 & \dots \\ 0 & c_1 & c_2 & c_3 - 1 & \dots \\ 0 & c_1 & c_2 & c_3 & \dots \\ 0 & c_1 & c_2 & c_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \right\|$$

has a unique solution. Show that this solution does not satisfy the dimension relations of this chapter, and describe the reduced game.

10. Show that  $\mathbf{x} = \langle \frac{1}{2}, \frac{1}{2} \rangle$ ,  $\mathbf{y} = \langle 1, 0, 0, \dots \rangle$  are the only optimal strategies for the infinite game

$$\mathbf{D} = \left\| \begin{array}{cccccc} d & 2d & \frac{1}{2} & 2d & \frac{1}{4} & \dots \\ d & 1 & 2d & \frac{1}{3} & 2d & \dots \end{array} \right\|,$$

where  $d \neq 0$ . Show that this example violates the dimension relations of this chapter.

11. Show that if  $\mathbf{H} = \|h_{ij}\|$  and  $\alpha > 0$ ,

$$\lim_{\alpha \rightarrow 0+} \frac{v(\mathbf{A} + \alpha \mathbf{H}) - v(\mathbf{A})}{\alpha} = \max_{x \in X^0(A)} \min_{y \in Y^0(A)} (\mathbf{x}, \mathbf{H} \mathbf{y}).$$

12. Show that

$$\left( \frac{\partial v}{\partial a_{ij}} \right)_+ = \left( \max_{x \in X^0} x_i \right) \left( \min_{y \in Y^0} y_j \right).$$

13. Derive the formula for

$$\left( \frac{\partial v}{\partial a_{ij}} \right)_-.$$

14. Let  $\mathbf{A}$  be a fixed  $2 \times 2$  game, and let  $X^0(\mathbf{A})$  and  $Y^0(\mathbf{A})$  be the spaces of optimal strategies for  $\mathbf{A}$ . Show that there exists a neighborhood  $U$  of  $\mathbf{A}$  such that if  $\mathbf{B}$  is in  $U$ , then

$$\dim X^0(\mathbf{B}) + \dim Y^0(\mathbf{B}) \leq \dim X^0(\mathbf{A}) + \dim Y^0(\mathbf{A}).$$

15. Consider the game  $\mathbf{A}_\sigma = \|e^{-\sigma(i-j)^2}\|$  ( $i, j = 1, \dots, n$ ), where  $\sigma$  is a positive parameter and  $n \leq 4$ . Find the solutions and the value of the game. Prove that for  $\sigma$  large the optimal strategies are unique.

16. Let  $\mathbf{A} = \|a_{ij}\|$  be an  $m \times n$  matrix in which  $a_{ij}$  are independent identically distributed random variables (see Section C.3) whose probability density is  $f(\xi)$ . Show that the probability that  $\mathbf{A}$  has a saddle point

is independent of  $f(\xi)$  and equal to

$$\frac{m! n!}{(m + n - 1)!}.$$

17. Let  $\mathbf{A}$  and  $\mathbf{B}$  denote matrices consisting of nonnegative components, and suppose that the game  $-\mathbf{A}$  and the game  $\mathbf{B}$  have a negative and a positive value, respectively. Prove that there exists a value  $p_0$  ( $0 < p_0 < 1$ ) and strategies  $\mathbf{x}^0$  and  $\mathbf{y}^0$  such that  $v(\mathbf{C}) = 0$ ,  $\mathbf{x}^0 \mathbf{C} \geq \mathbf{0}$ ,  $\mathbf{C} \mathbf{y}^0 \leq \mathbf{0}$ , and  $(\mathbf{x}^0, \mathbf{A} \mathbf{y}^0) > 0$ , where

$$\mathbf{C} = (1 - p_0)\mathbf{B} - p_0\mathbf{A}.$$

18. Verify the result of the preceding problem where

$$\mathbf{A} = \begin{vmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{vmatrix}, \quad \mathbf{B} = \begin{vmatrix} 0 & \cdots & 0 & n \\ 0 & \cdots & n-1 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{vmatrix},$$

by showing that  $p_0$  must be of the form  $k/(k + 1)$  for some integer  $k \leq n$ , and find  $\mathbf{x}^0$  and  $\mathbf{y}^0$  satisfying the desired properties.

19. Let  $\mathbf{A}$  be an  $n \times n$  matrix and let  $X^0$  and  $Y^0$  denote the sets of optimal strategies for Players I and II, respectively. Suppose that  $Y^0$  meets the boundary of  $Y$ . Show that there exists an optimal  $\mathbf{x}^0$  which uses at most  $n - 1$  indices.

20. Let  $W$  be a bounded, closed, convex polyhedral set lying in  $E^n$  which never touches the first orthant except possibly at the origin. Prove that there exists a vector  $\mathbf{v} = \langle v_1, v_2, \dots, v_n \rangle$  such that  $v_i > 0$  and  $(\mathbf{v}, \mathbf{w}) \leq 0$  for every  $\mathbf{w}$  in  $W$ .

### NOTES AND REFERENCES TO CHAPTER 3

§ 3.1. Theorems 3.1.1 and 3.1.2 were first presented by Bohnenblust, Karlin, and Shapley [36]. Gale and Sherman [100] have given an independent proof of Theorem 3.1.2, and Dresher and Karlin [70] have presented proofs of both theorems based on the Kakutani fixed-point theorem. Theorem 3.1.1 also emerges as a consequence of the study of “admissible” points of convex sets by Arrow, Barankin, and Blackwell [5]. Analogous structure theorems for linear programming systems have been obtained by Goldman and Tucker [111].

§ 3.3. Theorem 3.3.1, describing the relationship between sets of optimal strategies and the null space of the linear transformation defined by the matrix of the game, is due to Bohnenblust, Karlin, and Shapley [36].

**§ 3.4.** The converse of Theorem 3.1.2 is due to Bohnenblust, Karlin, and Shapley [36].

**§ 3.5.** Theorem 3.5.1 was first presented by Bohnenblust, Karlin, and Shapley [36] (see also Volume II Chapter 3).

**§ 3.6.** Problems 3 and 4 are due to Gale, Kuhn, and Tucker [98]. Problems 11, 12, and 13 are derived from Mills [181] and Gross [114]. Problem 16 is due to A. Goldman and Problem 17 to Thompson [231] (see also Kemeny, Morgenstern, and Thompson [142]).

## CHAPTER 4

### SOLUTIONS OF SOME DISCRETE GAMES

Although general numerical methods for solving games exist (see Chapter 6), often the special properties of a particular game can be used to determine its solutions quickly. Intuitive considerations frequently lead to a good initial guess, which can then be modified appropriately to yield the correct solutions.

Our analysis of the five examples in this chapter makes repeated use of the theory previously developed, notably dominance concepts, symmetries, and simple forms of probabilistic arguments. The examples have been selected to emphasize principles of game theory rather than the routine iterative procedures of solving games, and to illustrate methods and types of thinking that are applicable to a wide variety of problems.

The first two examples treated involve elementary matrix games. Both exhibit simple dominance structure and yield to elementary applications of the theory. The third example is the poker model introduced in Section 1.2. In describing the full game for the first two examples, we need only specify for each player the form of the pure strategies and the corresponding pay-off. In the third example it seems easier to deal directly with the full strategy space. The technique employed in this example is typical of a general method of solving poker models.

The fourth example is concerned with the strategies of two opposing political parties in an election campaign—a special case of the general question of how to distribute advertising effort. In the specific model discussed, each of the parties must decide in which of several areas to invest time and money in order to win votes. The explicit solution of the game exhibits some surprising consequences.

The final example involves a more elaborate matrix game. Its background was originally military, but it is presented here as a problem in collective bargaining.

The mathematics in this chapter is mostly elementary. Nevertheless, arguments are displayed in greater detail than is usual, in an effort to convey more completely the methodology and approach upon which the analysis of game problems depends.

**4.1 Colonel Blotto game.** The Colonel Blotto game is a general title for a class of tactical military games. The game considered in this section is one of the simplest examples.

Two opposing armies are advancing on two posts. The first army, under Colonel Blotto, consists of four regiments; the second, under Captain Kije, consists of three regiments. The army sending the most regiments to either post captures the post and all the regiments sent there by the other side, scoring one point for the captured post and one for each of the captured regiments. Colonel Blotto must decide how to deploy his forces in order to win the most points.

Colonel Blotto has five possible pure strategies. These strategies can be represented as pairs  $\{x, y\}$ , where  $x$  is the number of regiments sent to capture post 1 and  $y$  the number of regiments sent to capture post 2; they are  $\{4, 0\}$ ,  $\{0, 4\}$ ,  $\{3, 1\}$ ,  $\{1, 3\}$ , and  $\{2, 2\}$ . Similarly, the Captain's four possible strategies can be represented by the ordered pairs  $\{3, 0\}$ ,  $\{0, 3\}$ ,  $\{2, 1\}$ , and  $\{1, 2\}$ . The complete pay-off matrix is shown below.

		Kije			
		$a$ $\{3, 0\}$	$a$ $\{0, 3\}$	$b$ $\{2, 1\}$	$b$ $\{1, 2\}$
	$\alpha\{4, 0\}$	4	0	2	1
	$\alpha\{0, 4\}$	0	4	1	2
Blotto	$\beta\{3, 1\}$	1	-1	3	0
	$\beta\{1, 3\}$	-1	1	0	3
	$\gamma\{2, 2\}$	-2	-2	2	2

This game is symmetrical for Colonel Blotto in the strategies  $\{4, 0\}$  and  $\{0, 4\}$ , and also in the strategies  $\{3, 1\}$  and  $\{1, 3\}$ ; for Captain Kije strategies  $\{3, 0\}$ ,  $\{0, 3\}$ , and  $\{2, 1\}$ ,  $\{1, 2\}$  are symmetrical. It therefore seems reasonable to search for optimal strategies possessing the analogous symmetry property. Let  $a$  denote the probability of Captain Kije's using strategy  $\{3, 0\}$  (and  $\{0, 3\}$ ), and  $b$  the probability of his using strategy  $\{2, 1\}$  (and  $\{1, 2\}$ ). In order to make the mixed strategy  $\langle a, a, b, b \rangle$  optimal for Captain Kije, it is necessary and sufficient that

$$4a + 3b \leq v, \quad 3b \leq v, \quad -4a + 4b \leq v. \quad (4.1.1)$$

Clearly, the second inequality is implied by the first. Since  $2a + 2b = 1$ , the inequalities become  $4a + 3(\frac{1}{2} - a) \leq v$  or  $a + \frac{3}{2} \leq v$ , and  $-4a + 4(\frac{1}{2} - a) \leq v$  or  $-8a + 2 \leq v$ .

The values  $a^0 = \frac{1}{18}$ ,  $b^0 = \frac{4}{9}$ , and  $v = \frac{14}{9}$  provide a solution to (4.1.1) with strict inequality in the second relation.

If this does represent an optimal strategy, then Lemma 2.1.2 would

require  $\beta = 0$ , where  $\beta$  is the probability of Colonel Blotto's using  $\{3, 1\}$  or  $\{1, 3\}$ . The mixed strategy  $\langle \alpha, \alpha, 0, 0, \gamma \rangle$  will be optimal if and only if

$$4\alpha - 2\gamma \geq v = \frac{14}{9}, \quad 3\alpha + 2\gamma \geq v, \quad 2\alpha + \gamma = 1.$$

The values  $\alpha^0 = \frac{4}{9}$  and  $\gamma^0 = \frac{1}{9}$  satisfy these equations. By direct verification, we obtain that  $\langle \frac{4}{9}, \frac{4}{9}, 0, 0, \frac{1}{9} \rangle$  is indeed an optimal strategy for Colonel Blotto, that  $\langle \frac{1}{18}, \frac{1}{18}, \frac{4}{9}, \frac{4}{9} \rangle$  is an optimal strategy for Captain Kije, and that the value of the game is  $\frac{14}{9}$ . With the aid of Lemma 2.1.2, it can be immediately established that Colonel Blotto's (Player I's) optimal strategy is unique. However, Captain Kije's set of optimal strategies is not, as can be seen by applying the dimension relations of Theorem 3.1.2 to this example.

**4.2 Identification of Friend and Foe (I.F.F. game).** Aircraft approaching an air base must signal to the observer at the air base whether they are friendly or hostile. On the basis of the signal the observer must decide to which side, friend or enemy, the incoming aircraft belongs. The pay-off will be

- a if the observer identifies a friend as a friend,
- b if the observer identifies a friend as an enemy,
- c if the observer identifies an enemy as an enemy,
- d if the observer identifies an enemy as a friend,

where  $a > b, d$  and  $c > b, d$ .

Suppose that the observer and all friendly aircraft adopt the uncoded signal procedure whereby the friendly aircraft always signal "friend." Let  $p$  be the probability that an incoming plane is a friend. This is assumed to be known on the basis of previous experience.

The four possible strategies for the observer are  $I_1(F, E) = \{F, E\}$ ,  $I_2(F, E) = \{E, E\}$ ,  $I_3(F, E) = \{E, F\}$ ,  $I_4(F, E) = \{F, F\}$ , where, for example,  $I_2(F, E) = \{E, E\}$  means that the observer identifies a signal "friend" as an enemy and a signal "enemy" as an enemy.

The enemy has two possible strategies, one to signal "friend" and the other to signal "enemy." These are denoted by  $F$  and  $E$ , respectively.

If the observer uses  $I_3$  and the enemy strategy is  $E$ , the observer will identify a friend as an enemy and receive  $b$  with probability  $p$ , and he will identify an enemy as a friend and receive  $d$  with probability  $1 - p$ . The total expected pay-off is then  $pb + (1 - p)d$ . The eight expected pay-offs corresponding to the combinations of pure strategies are summarized in the following matrix:

$$\begin{array}{cc}
 E & F \\
 \hline
 I_1 & \left| \begin{array}{cc} pa + (1-p)c & pa + (1-p)d \end{array} \right| \\
 I_2 & \left| \begin{array}{cc} pb + (1-p)c & pb + (1-p)c \end{array} \right| \\
 I_3 & \left| \begin{array}{cc} pb + (1-p)d & pb + (1-p)c \end{array} \right| \\
 I_4 & \left| \begin{array}{cc} pa + (1-p)d & pa + (1-p)d \end{array} \right|
 \end{array} .$$

Since  $a > b, d$  and  $c > b, d$ , it is clear that  $I_1$  dominates  $I_4$ ,  $I_2$  dominates  $I_3$ , and in the smaller matrix  $E$  dominates  $F$ . Hence, the enemy will always signal "friend" and the observer will use  $I_1$  whenever  $pa + (1-p)d > pb + (1-p)c$ , or

$$p > \frac{c - d}{a + c - b - d}.$$

If the inequality is reversed, he will use  $I_2$ , and if equality holds, he can use either  $I_1$  or  $I_2$ .

The value of the game is  $pa + (1-p)d$  if

$$p > \frac{c - d}{a + c - b - d},$$

and  $pb + (1-p)c$  if

$$p < \frac{c - d}{a + c - b - d}.$$

A more realistic I.F.F. game would involve a scheme of coded signals. Suppose that the observer and friendly aircraft have a system worked out whereby on certain days the friendly aircraft signal "friend" and on others they signal "enemy." If on a particular day the signal "friend" is agreed upon, there are just two sensible strategies for the observer,  $I_1(F, E) = \{F, E\}$  and  $I_2(F, E) = \{E, E\}$ . There are actually two other pure strategies, but these are dominated by  $I_1$  and  $I_2$ , so we do not exhibit them. If the signal agreed upon is "enemy," the two sensible pure strategies for the observer are  $I_3(F, E) = \{E, E\}$  and  $I_4(F, E) = \{E, F\}$ .

The expected pay-off matrix is now

$$\begin{array}{cc}
 E & F \\
 \hline
 I_1 & \left| \begin{array}{cc} pa + (1-p)c & pa + (1-p)d \end{array} \right| \\
 I_2 & \left| \begin{array}{cc} pb + (1-p)c & pb + (1-p)c \end{array} \right| \\
 I_3 & \left| \begin{array}{cc} pb + (1-p)c & pb + (1-p)c \end{array} \right| \\
 I_4 & \left| \begin{array}{cc} pa + (1-p)d & pa + (1-p)c \end{array} \right|
 \end{array} .$$

Let  $y$  denote the probability with which the enemy employs strategy  $E$ .

Case 1.

$$pb + (1 - p)c < pa + (1 - p)\left(\frac{c + d}{2}\right).$$

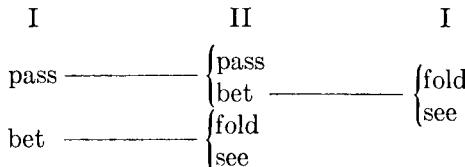
The techniques of Section 2.3 apply and yield readily that  $y^0 = \frac{1}{2}$  is the unique optimal strategy for the enemy. An optimal strategy for the observer consists of employing  $I_1$  and  $I_4$  each with probability  $\frac{1}{2}$ .

### Case 2.

$$pb + (1 - p)c > pa + (1 - p) \left( \frac{c + d}{2} \right).$$

We obtain a whole set of optimal  $y$  strategies for the enemy. Any combination of  $I_2$  and  $I_3$  is an optimal strategy for the observer.

**4.3 Poker game.** The poker game treated here was introduced previously in Section 1.2. The game is as follows. There are three different hands, 1, 2, and 3, such that  $1 < 2 < 3$  ( $i < j$  means that  $j$  defeats  $i$ ); each player receives one of the three hands, and antes the amount  $a$ ; a bet consists of adding an additional amount  $b$ . The possible sequences of moves are diagrammed below.



If both I and II pass, the hands are compared and the higher hand wins. If a player folds (passes after a bet), the other wins the pot. If a player sees (accepts the other's bet), the hands are compared and the higher wins.

The methods used in solving this game are representative of the analysis of parlor games in general. More important, the same techniques and procedures apply to tactical games of continuous firepower (see Chapters 8 and 9 of Volume II). Because of the importance of the methodology, we shall develop the arguments of the example in some detail, especially in describing the interplay which occurs in determining the maximizing and minimizing optimal strategies.

In this example a strategy for  $I$  is a set of three triplets  $\langle \alpha_i, \beta_i, \gamma_i \rangle$ , where if  $I$  gets hand  $i$ ,  $\alpha_i$  is the probability that he bets immediately,  $\beta_i$  is the probability that he passes on the first round and subsequently sees if the occasion arises, and  $\gamma_i$  is the probability that he passes on the first round and subsequently folds if the occasion arises. The quantities  $\alpha_i$ ,  $\beta_i$ , and  $\gamma_i$  are subject to the following restrictions:

$$0 \leq \alpha_i, \beta_i, \gamma_i \leq 1 \quad \text{and} \quad \alpha_i + \beta_i + \gamma_i = 1.$$

A strategy for II is a set of three pairs  $\langle x_i, y_i \rangle$ , where if II gets hand  $i$ ,  $x_i$  is the probability of his betting after a pass, and  $y_i$  is the probability of his seeing after a bet. The quantities  $x_i$  and  $y_i$  are subject to the restriction  $0 \leq x_i, y_i \leq 1$ .

Let

$$L(i, j) = \begin{cases} 1 & \text{if } i > j, \\ -1 & \text{if } i < j. \end{cases}$$

Suppose that I receives hand  $i$ , thus using strategy  $\langle \alpha_i, \beta_i, \gamma_i \rangle$ , and suppose that II receives hand  $j$ , thus using strategy  $\langle x_j, y_j \rangle$ . The probability that both I and II will pass is  $(\beta_i + \gamma_i)(1 - x_j)$ ; the pay-off in this case would be  $aL(i, j)$ . The expected pay-off for a pass by I, a bet by II, and a bet (see) by I is  $(a + b)\beta_i x_j L(i, j)$ ; and the expected pay-off for a pass by I, a bet by II, and a pass (fold) by I is  $-a\gamma_i x_j$ . The various probabilities and pay-offs if I bets initially are also readily computed. The full expected pay-off function is then seen to be

$$\begin{aligned} 3! K(\alpha, \beta, \gamma; x, y) = & a \sum_{i \neq j} (\beta_i + \gamma_i)(1 - x_j)L(i, j) \\ & + (a + b) \sum_{i \neq j} \beta_i x_j L(i, j) \\ & - a \sum_{i \neq j} \gamma_i x_j + a \sum_{i \neq j} \alpha_i (1 - y_j) \\ & + (a + b) \sum_{i \neq j} \alpha_i y_j L(i, j). \end{aligned}$$

To obtain a solution to the game we need to find two strategies  $\langle x^0, y^0 \rangle$  and  $\langle \alpha^0, \beta^0, \gamma^0 \rangle$  such that

$$K(\alpha^0, \beta^0, \gamma^0; x^0, y^0) \leq K(\alpha^0, \beta^0, \gamma^0; x, y) \quad (4.3.1)$$

for all  $\langle x, y \rangle$  and

$$K(\alpha^0, \beta^0, \gamma^0; x^0, y^0) \geq K(\alpha, \beta, \gamma; x^0, y^0) \quad (4.3.2)$$

for all  $\langle \alpha, \beta, \gamma \rangle$ . In other words, what we want to do is minimize the function  $K(\alpha^0, \beta^0, \gamma^0; x, y)$  with respect to  $\langle x, y \rangle$ , and we hope that the minimum is achieved for  $x^0, y^0$ . At the same time we want to maximize  $K(\alpha, \beta, \gamma; x^0, y^0)$  in the first set of variables, and we hope that the maximum is achieved for  $\langle \alpha^0, \beta^0, \gamma^0 \rangle$ . This leads to consistency conditions for the optimal strategies  $\alpha^0, \beta^0, \gamma^0$  and  $x^0, y^0$ .

Suppose that the coefficient of  $x_i$  in the pay-off function  $K(\alpha^0, \beta^0, \gamma^0; x, y)$  is positive; then, since II wants to minimize, he will require  $x_i^0 = 0$ . On the other hand, if the coefficient is negative, he will require  $x_i^0 = 1$ . In

the event the coefficient is zero, no immediate constraint is imposed on  $x_i^0$ . Similar considerations hold for  $y_i^0$ .

The case for I is not quite so simple, because of the added restriction that  $\alpha_i + \beta_i + \gamma_i = 1$ . What Player I must do in order to maximize  $K(\alpha, \beta, \gamma; x^0, y^0)$  is to set that variable ( $\alpha^0, \beta^0$ , or  $\gamma^0$ ) equal to 1 which has the largest coefficient. If the two largest coefficients are equal, or all three coefficients are equal, then no restriction on the corresponding  $\alpha^0, \beta^0$ , and  $\gamma^0$  exists, except that they add up to 1.

These last two paragraphs outline the reasoning with which we shall attempt to solve this poker game.

The variables with their coefficients in the pay-off function are listed below in columns. These expressions were obtained simply by collecting the coefficients from wherever the variable appeared in the function and adding them.

$$\begin{aligned}
 3! K(\alpha^0, \beta^0, \gamma^0; x, y) = & \quad 3! K(\alpha, \beta, \gamma; x^0, y^0) = \\
 y_3[-(\alpha_1^0 + \alpha_2^0)(2a + b)] & \alpha_3[2a + b(y_1^0 + y_2^0)] \\
 y_2[-2a\alpha_1^0 - b(\alpha_1^0 - \alpha_3^0)] & \beta_3[2a + b(x_1^0 + x_2^0)] \\
 y_1[b(\alpha_2^0 + \alpha_3^0)] & \gamma_3[2a - 2a(x_1^0 + x_2^0)] \\
 x_3[-b(\beta_1^0 + \beta_2^0)] & \alpha_2[2a + by_1^0 - (2a + b)y_3^0] \\
 x_2[-2a\gamma_3^0 - b(\beta_1^0 - \beta_3^0)] & \beta_2[b(x_1^0 - x_3^0)] \\
 x_1[-2a(\gamma_2^0 + \gamma_3^0) + b(\beta_2^0 + \beta_3^0)] & \gamma_2[-2ax_1^0] \\
 + \text{terms independent of } x_i \text{ and } y_j; & \alpha_1[2a - (2a + b)(y_2^0 + y_3^0)] \\
 & \beta_1[-2a - b(x_2^0 + x_3^0)] \\
 & \gamma_1[-2a].
 \end{aligned}$$

Since the coefficient of  $y_3$  is very likely to be negative, unless  $\alpha_1^0 + \alpha_2^0 = 0$ , a good guess for  $y_3^0$  is 1. Similarly, since the coefficient of  $y_1$  is probably positive, a good guess for  $y_1^0$  is 0, and we shall guess that  $x_3^0 = 1$ , since  $-b(\beta_1 + \beta_2)$  is always nonpositive.

The coefficient of  $\gamma_3$  is most likely to be smaller than the coefficients of  $\alpha_3$  and  $\beta_3$ , so that probably  $\gamma_3^0 = 0$ . Similarly, by comparing the coefficient of  $\beta_1$  with that of  $\gamma_1$ , we see that  $\beta_1^0$  should equal 0.

Our set of variables and coefficients can now be reduced, by eliminating those variables which have been estimated and substituting their values in the remaining coefficients. The resulting coefficients are

$$y_2[-2a\alpha_1^0 - b(\alpha_1^0 - \alpha_3^0)] + x_2[b\beta_3^0] + x_1[-2a\gamma_2^0 + b(\beta_2^0 + \beta_3^0)]$$

and

$$\begin{aligned}\alpha_3[2a + by_2^0] + \beta_3[2a + b(x_1^0 + x_2^0)] + \alpha_2[-b] + \beta_2[b(x_1^0 - 1)] \\ + \gamma_2[-2ax_1^0] + \alpha_1[-b - (2a + b)y_2^0] + \gamma_1[-2a].\end{aligned}$$

No further implications about the coefficients can be deduced without imposing some assumptions on the relative sizes of  $a$  and  $b$ . The coefficients of  $\alpha_2$ ,  $\beta_2$ , and  $\gamma_2$  depend only on the size of  $x_1$ , so we shall first attempt to ascertain their weights.

If  $2a > b$ , then  $x_1^0 < b/(2a + b)$  implies that  $\beta_2^0 = 0$ ,  $\alpha_2^0 = 0$ ,  $\gamma_2^0 = 1$ , which in turn implies that the coefficient of  $x_1$ ,  $[-2a + b\beta_3^0]$ , is less than zero. This leads to the conclusion  $x_1^0 = 1$ , a contradiction.

Thus if  $2a > b$ , then  $x_1^0 \geq b/(2a + b)$ . Suppose  $x_1^0 > b/(2a + b)$ ; then  $\beta_2^0 = 1$ , which implies that the coefficient of  $x_1^0$  is positive and hence that  $x_1^0 = 0$ , a contradiction. Therefore  $x_1^0 = b/(2a + b)$  and  $\alpha_2^0 = 0$ .

Looking at  $y_2$  and  $\alpha_1$  and supposing  $y_2^0 < (2a - b)/(2a + b)$ , we obtain

$$y_2^0 < \frac{2a - b}{2a + b} \rightarrow \alpha_1 = 1 \rightarrow \text{the coefficient of } y_2 \text{ is negative} \rightarrow y_2^0 = 1,$$

a contradiction. Furthermore,

$$y_2^0 > \frac{2a - b}{2a + b} \rightarrow \alpha_1 = 0 \rightarrow \text{the coefficient of } y_2 \text{ is nonnegative},$$

which gives  $\alpha_3 = 0$ ,  $\beta_3 = 1$ , from which we infer that  $x_2^0 = 0$  and

$$x_1^0 = \frac{b}{2a + b} > y_2^0 > \frac{2a - b}{2a + b},$$

so that if we assume  $a \geq b$ , we have a contradiction.

*Case 1.*  $a > b$ . (The case  $b > a$  is left as an exercise; see Problems 4-7.)

In this case we have seen that

$$x_1^0 = \frac{b}{2a + b}, \quad \alpha_2^0 = 0, \quad \text{and} \quad y_2^0 = \frac{2a - b}{2a + b}.$$

Now  $\beta_3^0 > 0 \rightarrow x_2^0 = 0$ . But then  $[2a + b(x_1^0 + x_2^0)] < [2a + by_2^0] \rightarrow \beta_3^0 = 0$ , a contradiction. Therefore  $\beta_3^0 = 0$  and  $\alpha_3^0 = 1$ . The equations have now been reduced to

$$y_2[-(2a + b)\alpha_1^0 + b] + x_1[-2a\gamma_2^0 + \beta_2^0b] + x_2 \cdot 0$$

and

$$\beta_2\left[\frac{-2ab}{2a + b}\right] + \gamma_2\left[\frac{-2ab}{2a + b}\right] + \alpha_1[-2a] + \gamma_1[-2a].$$

Since  $0 < x_1^0 < 1$ , the coefficient of  $x_1$  must be 0; thus  $\beta_2^0 = (2a/b)\gamma_2^0$ . But  $\beta_2 + \gamma_2 = 1$ . Hence

$$\beta_2^0 = \frac{2a}{2a+b} \quad \text{and} \quad \gamma_2^0 = \frac{b}{2a+b}.$$

Similarly, since  $0 < y_2^0 < 1$ , the coefficient of  $y_2$  must be 0. Hence

$$\alpha_1^0 = \frac{b}{2a+b} \quad \text{and} \quad \gamma_1^0 = \frac{2a}{2a+b}.$$

This leaves only  $x_2$  undetermined. Since its coefficient is 0, no restriction emerges directly, but since  $\beta_3^0 = 0$ , we must have  $x_1^0 + x_2^0 \leq y_2^0$ . Hence

$$x_2^0 \leq \frac{2(a-b)}{2a+b}.$$

The set of strategies conjectured to be optimal for Case 1 may be summarized as follows:

$$\begin{aligned} x_3^0 &= 1, & y_3^0 &= 1, \\ x_2^0 &\leq \frac{2(a-b)}{2a+b}, & y_2^0 &= \frac{2a-b}{2a+b}, \\ x_1^0 &= \frac{b}{2a+b}, & y_1^0 &= 0, \\ \alpha_3^0 &= 1, & \beta_3^0 &= 0, & \gamma_3^0 &= 0, \\ \alpha_2^0 &= 0, & \beta_2^0 &= \frac{2a}{2a+b}, & \gamma_2^0 &= \frac{b}{2a+b}, \\ \alpha_1^0 &= \frac{b}{2a+b}, & \beta_1^0 &= 0, & \gamma_1^0 &= \frac{2a}{2a+b}. \end{aligned} \tag{4.3.3}$$

The functions  $K(\alpha, \beta, \gamma; x^0, y^0)$  and  $K(\alpha^0, \beta^0, \gamma^0; x, y)$  will now be maximized and minimized, respectively, with the values of (4.3.3) used in the coefficients. If the maximizing and minimizing variables obtained coincide with those above, then the relations of (4.3.1) and (4.3.2) are satisfied and the strategies exhibited are indeed optimal. The analysis is tabulated:

$$y_3 \left[ \left( \frac{-b}{2a+b} \right) (2a+b) \right] \quad y_3 = 1 \text{ minimizes}$$

$$y_2 \left[ -2a \left( \frac{b}{2a+b} \right) - b \left( \frac{-2a}{2a+b} \right) \right] \quad \text{anything minimizes, in particular} \\ y_2 = \frac{2a-b}{2a+b}$$

$y_1[b]$	$y_1 = 0$ minimizes
$x_3 \left[ \frac{-2ab}{2a+b} \right]$	$x_3 = 1$ minimizes
$x_2[0]$	anything minimizes, in particular
$x_1 \left[ \frac{-2ab}{2a+b} + \frac{2ab}{2a+b} \right]$	anything minimizes, in particular
$\alpha_3[2a + b(y_2^0)]$ $\beta_3[2a + b(x_1^0 + x_2^0)]$ $\gamma_3[2a - 2a(x_1^0 + x_2^0)]$	$\alpha_3 = 1$ maximizes $\beta_3 = 0$ maximizes $\gamma_3 = 0$ maximizes
$\alpha_2[-b]$ $\beta_2 \left[ -\frac{2ab}{2a+b} \right]$ $\gamma_2 \left[ -\frac{2ab}{2a+b} \right]$	$\alpha_2 = 0$ maximizes $\beta_2 + \gamma_2 = 1$ maximizes, in particular $\beta_2 = \frac{2a}{2a+b}, \gamma_2 = \frac{b}{2a+b}$
$\alpha_1 \left[ 2a - (2a+b) \left( \frac{4a}{2a+b} \right) \right]$ $\gamma_1[-2a]$ $\beta_1[-2a - b(x_2^0 + 1)]$	$\alpha_1 + \gamma_1 = 1$ maximizes, in particular $\alpha_1 = \frac{b}{2a+b}, \gamma_1 = \frac{2a}{2a+b}$ $\beta_1 = 0$ maximizes

The consistency requirements are satisfied, and thus the strategies are optimal. The value of the game is

$$-\frac{1}{6} \frac{b^2}{2a+b}.$$

A further examination of the reasoning used to derive the form of the solutions of Case 1 shows in fact that we have found them all.

*Case 2.*  $a = b$ . The solution for this case is summarized without proof.

$$\begin{array}{lll}
 x_3^0 = 1 & & y_3^0 = 1 \\
 x_2^0 = 0 & & y_2^0 = \frac{1}{3} \\
 x_1^0 = \frac{1}{3} & & y_1^0 = 0 \\
 \alpha_2^0 = 0 & -3\alpha_1^0 + \alpha_3^0 = 0 & \beta_3^0 + \alpha_3^0 = 1 \\
 \beta_1^0 = 0 & -2\gamma_2^0 + \beta_2^0 + \beta_3^0 = 0 & \beta_2^0 + \gamma_2^0 = 1 \\
 \gamma_3^0 = 0 & & \alpha_1^0 + \gamma_1^0 = 1
 \end{array}$$

The value of the game is  $-a/6.3$ .

The two solutions obtained above exhibit several interesting features. For example, Player I never bets the first round with hand 2. We observe in this connection that for  $a = b$ , Player II also does not bet with hand 2 if Player I has originally passed. This is not necessarily true for  $a > b$ .

Where  $a = b$ , Player I sometimes “sandbags,” i.e., passes with a high hand on the first round and sees any subsequent bet.

The need for random bluffing comes in as a natural part of the optimal strategies. This is expressed by the fact that  $x_1^0, \alpha_1^0$ , and  $\gamma_1^0$  are all strictly positive. The frequency with which bluffing is desirable depends on the ratio of  $b$  to  $a$ .

**4.4 An advertising example.** Two political parties are engaged in an election campaign. There are  $n$  localities, labeled 1 through  $n$ , which normally vote for the candidate of Party II. In terms of number of voters, these localities rank as follows:  $a_1 > a_2 > \dots > a_n > 0$ . Party I has announced that it intends to invade one of these areas in an attempt to win votes. Party II will try to anticipate Party I's campaign with counter-propaganda. Facilities and finances are limited, so that each of the parties can put its effort into only one area. We assume that the campaign is such that if Party I invades area  $j$ , the value to him is  $a_j$  if the area is not defended by Party II, and  $pa_j$  ( $0 \leq p < 1$ ) if the area is defended. The coefficient  $p$  may be considered as a measure of the rate of effectiveness of the party orators, etc.

Each party has  $n$  pure strategies. The pay-off matrix is

$$\mathbf{A} = \begin{vmatrix} pa_1 & a_1 & \dots & a_1 \\ a_2 & pa_2 & \dots & a_2 \\ a_3 & a_3 & \dots & a_3 \\ \vdots & \vdots & & \vdots \\ a_n & a_n & \dots & pa_n \end{vmatrix}$$

Intuitively it seems likely that Party I would want to invade the more valuable areas and that Party II would defend the more valuable areas. We try to determine an optimal mixed strategy with the first  $r$  rows relevant, keeping  $r$  as a parameter to be determined later. Suppose that  $\mathbf{x}^0 = \langle x_1^0, x_2^0, \dots, x_r^0, 0, \dots, 0 \rangle$  denotes an optimal strategy. Assume also, temporarily, that the first  $r$  pure strategies belong to an optimal  $\mathbf{y}$  strategy. Then, by virtue of Lemma 2.1.2,

$$\sum_{i=1}^r a_i x_i^0 - (1 - p)a_k x_k^0 = v \quad (1 \leq k \leq r). \quad (4.4.1)$$

Hence  $a_k x_k^0 = a_1 x_1^0$  for  $2 \leq k \leq r$ , and by the normalization condition we obtain

$$x_k^0 = \frac{1/a_k}{\sum_{i=1}^r 1/a_i} \quad \text{and} \quad v = \frac{r+p-1}{\sum_{i=1}^r 1/a_i}. \quad (4.4.2)$$

It follows easily from (4.4.1) and (4.4.2) that

$$\mathbf{x}^0 \mathbf{A} \geq \mathbf{v},$$

with strict inequality for all components  $j$  where  $j > r$ . By Lemma 2.1.2 all optimal  $\mathbf{y}$  strategies must concentrate in the first  $r$  columns (since we do not know that  $\mathbf{x}^0$  is actually optimal for Party I, the analysis should be regarded as a motivated searching for the solution). Assuming that  $\mathbf{y}^0 = \langle y_1^0, y_2^0, \dots, y_r^0, 0, \dots, 0 \rangle$ , we determine  $y_i^0$  from

$$\frac{p+r-1}{\sum_{i=1}^r 1/a_i} = p a_k y_k^0 + a_k (1 - y_k^0) \quad (1 \leq k \leq r)$$

or

$$y_k^0 = \frac{1}{1-p} \left[ 1 - \frac{(p+r-1)/a_k}{\sum_{i=1}^r 1/a_i} \right] \quad (1 \leq k \leq r), \quad (4.4.3)$$

which is nonnegative if and only if

$$\frac{1}{r+p-1} \geq \frac{1/a_k}{\sum_{i=1}^r 1/a_i} \quad (1 \leq k \leq r). \quad (4.4.4)$$

Since  $a_k$  strictly decreases, (4.4.4) is equivalent to

$$\frac{1}{r+p-1} \geq \frac{1/a_r}{\sum_{i=1}^r 1/a_i}.$$

The strategy  $\mathbf{y}^0$  will satisfy  $\mathbf{A}\mathbf{y}^0 \leq \mathbf{v}$  if and only if

$$\frac{1/a_{r+1}}{\sum_{i=1}^r 1/a_i} \geq \frac{1}{r+p-1}. \quad (4.4.5)$$

In order to fulfill the conditions of both (4.4.4) and (4.4.5), we choose the value  $r$  as the smallest positive integer less than or equal to  $n$  satisfying

$$\frac{1/a_{r+1}}{\sum_{i=1}^r 1/a_i} > \frac{1}{r+p-1} \geq \frac{1/a_r}{\sum_{i=1}^r 1/a_i}. \quad (4.4.6)$$

If no  $r$  can be found obeying these inequalities, we define  $r = n$ . This case can occur only when

$$\frac{1/a_n}{\sum_{i=1}^n 1/a_i} \leq \frac{1}{n+p-1}. \quad (4.4.7)$$

The proof of this assertion is as follows. Note first that if

$$\frac{1}{r+p-1} \geq \frac{1/a_{r+1}}{\sum_{i=1}^r 1/a_i},$$

then

$$\frac{1}{r+p} \geq \frac{1/a_{r+1}}{\sum_{i=1}^{r+1} 1/a_i}.$$

(The first inequality is the same as

$$r+p-1 \leq \frac{\sum_{i=1}^r 1/a_i}{1/a_{r+1}},$$

and hence

$$r+p \leq \frac{\sum_{i=1}^r 1/a_i}{1/a_{r+1}} + \frac{1/a_{r+1}}{1/a_{r+1}},$$

which is the second inequality.)

Next, observe that the right-hand inequality of (4.4.6) is valid for  $r = 1$ . In view of what has just been shown, we may conclude that (4.4.6) can fail to hold for all  $r$  only when (4.4.7) obtains.

The optimal strategies are now explicit. We need only choose  $r$  according to (4.4.6) and then define  $\mathbf{x}^0$  and  $\mathbf{y}^0$  by means of (4.4.2) and (4.4.3), respectively. That these are indeed solutions has been established in the course of the preceding analysis. Except if  $a_k$  and  $p$  take on special values [for which case the rare equality for the determination of  $r$  according to (4.4.6) obtains], the optimal strategies are unique.

We find that both parties concentrate attention on the same  $r$  attractive localities, and that they differ solely in having different probabilities of selecting a given locality from among those under consideration.

**\*4.5 A bargaining example.** The game we shall analyze in this section is given an interpretation of collective bargaining. This interpretation may seem to an extent contrived, since labor-management disputes do not involve a full conflict of interests, but it nonetheless serves to suggest possible applications of game-theory reasoning to industrial situations.

Our example was originally a game of strategy concerned with attack and defense of a hidden object. A defender has an object of great value which he may conceal in any one of many containers; the attacker makes a series of attempts to destroy the object by destroying the containers. By way of illustration, suppose that an A-bomb is to be carried in one of two identical bombers, called  $P$  (protected) and  $F$  (flank). They fly in forma-

---

\*Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

tion with  $P$  behind  $F$ , so that a hostile fighter wishing to attack  $P$  must pass through  $F$ 's field of fire and run a risk of being shot down before being able to close on its target. Once it has engaged its target, however, the fighter can destroy it with probability  $\beta$ . The defender (Player II) chooses  $P$  or  $F$  to carry the bomb. The attacker (Player I) chooses an ordered set of  $n$  members  $PFFP \dots P$ , where the individual members represent his target choice on each pass provided all previous passes have failed. The integer  $n$  represents the number of allowable attacks the fighter can make.

We now revert to the bargaining interpretation, keeping the notation  $P$  and  $F$  as a reminder of the attack-and-defense implication of the model.

A union and the management of a certain firm are at odds over two issues,  $F$  and  $P$ . Under all circumstances let  $\beta$  be the probability that the union can settle the issue  $F$ , negatively or positively. If  $F$  is settled positively (in favor of the union), the game is over; i.e., no further agreements can be reached. If  $F$  is settled negatively, let  $\beta$  be the probability of the union's settling  $P$ . If  $F$  is not settled one way or the other, let  $\gamma < \beta$  be the probability of the union's settling  $P$ . In other words, if  $F$  has not previously been settled,  $P$  is harder to settle. (In the attack-and-defense model, it is harder to destroy  $P$  when  $F$  is intact than when  $F$  has been destroyed.)

In any opportunity to bargain, the union must decide which issue to attempt to settle. We shall assume that management *a priori* is inclined to make concessions on only one of the issues, if any at all. Management has two pure strategies; i.e., it must decide which of the two issues it will make concessions on if necessary. (All this in reality is decided ahead of time with a proportion of sureness.) We shall also suppose that the union knows that progress on at most one of the issues is possible. If management is prepared to yield on one of the issues, say  $P$ , then we assume that the probability of the union's ascertaining the hopelessness of settling  $F$  is still  $\beta$ . Finally, we assume that if the hopelessness of settling  $F$  becomes known to the union, the next attempt at negotiation will be directed at issue  $P$ . The situation is not symmetrical. If management has decided to yield on  $F$  if necessary, the probability of the union's determining the impossibility of negotiating  $P$  is  $\gamma$ . The union also has two pure strategies; i.e., it must decide which of the two issues to bargain. Player I is identified with the union and Player II with management.

In the case of only one opportunity for bargaining (opportunity to bargain may correspond to a meeting every three months between union and management representatives), the formal description of the bargaining problem is as follows. Two strategies are available to Player II: (i) to favor yielding on issue  $P$ ; (ii) to favor yielding on issue  $F$ . Player I also has two possible strategies: (i) to bargain  $P$ ; (ii) to bargain  $F$ . The pay-off to Player I is taken as equal to the probability that the issue will be settled positively for the union.

	<u>Yield on <math>P</math></u>	<u>Yield on <math>F</math></u>
Bargain $P$	$\gamma$	0
Bargain $F$	0	$\beta$

The value of the game is  $v = \beta\gamma/(\beta + \gamma)$ , and Player I has the unique optimal mixed strategy

$$\text{play } P \text{ with prob } \frac{\beta}{\beta + \gamma},$$

$$\text{play } F \text{ with prob } \frac{\gamma}{\beta + \gamma}.$$

Player II has the same optimal strategy.

We now generalize the preceding by allowing Player I two opportunities to bargain. Player I now has four pure strategies, and the pay-off kernel becomes

	$P$	$F$
$P, P$	$2\gamma - \gamma^2$	$\beta\gamma$
$P, F$	$\gamma$	$\beta$
$F, P$	$\beta^2 + (1 - \beta)\gamma$	$\beta$
$F, F$	$\beta^2$	$2\beta - \beta^2$

The first letter of a pure strategy for Player I indicates the issue to be bargained at the first opportunity, and the second letter the issue to be bargained at the second opportunity if the first session has resulted in no decision or a negative decision. The pay-off kernel is the expected value for Player I or, equivalently, the probability of success according to the strategies chosen by Players I and II. It is understood that if Player I can completely settle either issue, positively or negatively, then that issue may be considered closed.

We now show the derivation of a typical element in the preceding matrix:  $\beta^2 + (1 - \beta)\gamma$ , appearing in the third row of the first column. Player I's strategy is  $F, P$  and Player II's strategy is  $P$ . There are two possible outcomes if Player I bargains  $F$ : (a)  $F$  is settled negatively (with prob  $\beta$ ) and the probability now of settling  $P$  is  $\beta$ ; or (b)  $F$  is not settled (with prob  $1 - \beta$ ), and the probability of settling  $P$  is  $\gamma$ . The total probability is thus

$$\beta^2 + (1 - \beta)\gamma.$$

We next investigate the solutions of the game. Note that since  $\gamma < \beta$ , the second row is dominated by the third row. Some elementary manipula-

tions show that the fourth row, although not dominated, is not involved in any optimal strategy for Player I. The solution is as follows:

*Case 1.* If  $\beta^2 - \beta\gamma + \gamma^2 - \gamma \leq 0$ , then the optimal strategy for Player I is

$$F, P \text{ with prob } \frac{\gamma(1 - \beta) + \gamma(1 - \gamma)}{\beta(1 - \beta) + \gamma(1 - \gamma)},$$

$$P, P \text{ with prob } \frac{\beta(1 - \beta) - \gamma(1 - \beta)}{\beta(1 - \beta) + \gamma(1 - \gamma)};$$

the optimal strategy for Player II is

$$\text{favor yielding on } P \text{ with prob } \frac{\beta(1 - \gamma)}{\beta(1 - \beta) + \gamma(1 - \gamma)},$$

$$\text{favor yielding on } F \text{ with prob } \frac{\beta(\gamma - \beta) + \gamma(1 - \gamma)}{\beta(1 - \beta) + \gamma(1 - \gamma)};$$

and the value is

$$v = \beta\gamma \frac{\beta(\gamma - \beta) + 2(1 - \gamma)}{\beta(1 - \beta) + \gamma(1 - \gamma)}.$$

*Case 2.* If  $\beta^2 - \beta\gamma + \gamma^2 - \gamma \geq 0$ , then a saddle point exists. Player I chooses  $F, P$ , and Player II favors  $P$ . The value of the game is

$$v = \beta^2 - \beta\gamma + \gamma.$$

Our final extension of this model is to allow Player I  $n$  bargaining opportunities. A strategy is described by an ordered succession of  $n$  letters  $F$  or  $P$  which determines the order of bargaining. We repeat that once an issue is fully settled negatively, the other issue is bargained at the next opportunity.

In studying this general game, we note first that every strategy for Player I in which  $F$  is involved  $r$  times is dominated by

$$\overbrace{F \dots F}^r \overbrace{P \dots P}^{n-r}.$$

(We have seen an instance of this in the last matrix, where the second row was dominated by the third row.) In other words, if  $F$  is to be bargained  $r$  times by Player I, then this issue should be bargained first. The argument of the case of two bargaining opportunities shows that  $F \dots FP \dots$  dominates the similar strategy with the  $FP$  interchanged. Iterating this establishes the assertion. Consequently, strategies for Player I are described by an integer  $r$  specifying the number of times  $F$  is discussed. We

now determine the terms in the pay-off matrix  $a(r, P)$  and  $a(r, F)$ , the first of which corresponds to the strategy  $r$  for I and  $P$  for II.

$$\begin{aligned} a(r, P) &= 1 - \text{prob}\{\text{I unable to settle } F \text{ in the first } r \text{ bargaining opportunities}\} \cdot \text{prob}\{\text{I unsuccessful in bargaining } P|\text{I unable to settle } F\} - \sum_{k=1}^r \text{prob}\{\text{I able to settle } F \text{ at the } k\text{th opportunity}\} \cdot \text{prob}\{\text{I unsuccessful in bargaining } P \text{ in the remaining } n-k \text{ trials, given that } F \text{ was settled negatively at the } k\text{th discussion}\} \\ &= 1 - (1 - \beta)^r (1 - \gamma)^{n-r} - r\beta(1 - \beta)^{n-1}. \end{aligned}$$

In a similar manner we find

$$\begin{aligned} a(r, F) &= 1 - (1 - \beta)^r (1 - \gamma)^{n-r} \\ &\quad - \gamma(1 - \beta)^r \frac{[(1 - \gamma)^{n-r} - (1 - \beta)^{n-r}]}{\beta - \gamma}. \end{aligned}$$

The game is solved by letting  $r$  be continuous and making use of the concavity of  $a(r, F)$  and  $a(r, P)$  in  $r$ . Let  $r_0$  be the solution of  $da(r, P)/dr = 0$  and let  $R = (1 - \gamma)/(1 - \beta)$ . Then

$$r_0 = n - \frac{\ln [\beta/(1 - \beta)] - \ln \ln R}{\ln R},$$

provided the right side is positive; otherwise set  $r_0 = 0$ . If  $a(r_0, P) \leq a(r_0, F)$ , then I has a pure optimal strategy  $r_0$  and II has a pure optimal strategy  $P$ . If  $a(r_0, P) \geq a(r_0, F)$ , then I has a pure optimal strategy  $r$  given by the solution of the equation

$$1 + \frac{r\beta(\beta - \gamma)}{\gamma(1 - \beta)} = \left(\frac{1 - \gamma}{1 - \beta}\right)^{n-r},$$

or approximately  $r = n\gamma/(\beta + \gamma)$  when  $n$  is large and  $\beta$  is close to  $\gamma$ . II has an optimal strategy which is a mixture of  $P$  and  $F$ . For the original discrete game, define the integer  $r_1 = [r_0]$  and solve the  $2 \times 2$  matrix game

$$\begin{vmatrix} a(r_1, P) & a(r_1, F) \\ a(r_1 + 1, P) & a(r_1 + 1, F) \end{vmatrix}.$$

If  $a(r_1, P) \leq a(r_1, F)$ , we have a pure strategy solution; otherwise the solution is mixed. This technique will be further discussed in a later chapter, in connection with convex games. (See also Section 2.7, Problem 22.)

#### 4.6 Problems.

1. Determine all the solutions of the Colonel Blotto problem of Section 4.1.
2. Let  $\mathbf{A}$  be the  $n \times n$  matrix game with

$$a_{ij} = \begin{cases} +1 & \text{if } i \neq j, \\ -1 & \text{if } i = j. \end{cases}$$

Use the symmetry of the game to find the optimal strategies and the value of this game.

3. Consider the I.F.F. coded game with one signal where a signal is an  $n$ -tuple whose components consist of the symbols  $F$  or  $E$ . Using the parameters  $a, b, c, d$  and  $p$  of Section 4.2, show that the optimal strategy for the friend and observer is to code the signals and select the particular code to be used at random from among all possible codes. Prove that the enemy's optimal strategy is to select a signal at random from among all possible signals. Prove also that if

$$p > \frac{c - d}{c - d + 2^n(a - b)},$$

where  $n$  is the number of components to the signal, the observer identifies only the friend's signal as coming from a friend, and that if the opposite inequality holds, the observer identifies every signal as an enemy's. Find the value of this game.

4. Verify that if  $a = b$ , the poker model of Section 4.3 has the solution described at the end of that section.

5. Verify that if  $a < b < 2a$ , the poker model of Section 4.3 has the following solution:

$$x_3^0 = 1 \quad y_3^0 = 1$$

$$x_2^0 = 0 \quad \frac{2a - b}{2a + b} \leq y_2^0 \leq \frac{b}{2a + b}$$

$$x_1^0 = \frac{b}{2a + b} \quad y_1^0 = 0$$

$$\alpha_3^0 = 0 \quad \beta_3^0 = 1 \quad \gamma_3^0 = 0$$

$$\alpha_2^0 = 0 \quad \beta_2^0 = \frac{2a - b}{2a + b} \quad \gamma_2^0 = \frac{2b}{2a + b}$$

$$\alpha_1^0 = 0 \quad \beta_1^0 = 0 \quad \gamma_1^0 = 1$$

6. Find the solution of the poker model of Section 4.3, assuming that  $2a = b$ .

7. Find the solution of the poker model of Section 4.3, assuming that  $2a < b$ .

8. Player I has three stones marked 1, 2, 3, and Player II has three stones marked  $-1, -2, -3$ . The players have a  $3 \times 3$  checkerboard, on which the squares  $(1, 3), (2, 2)$ , and  $(3, 1)$  are marked 0. To start the game, Player I places one of his stones in any one of the unmarked squares; Player II next places a stone in one of the remaining squares, and so on until the board is full. They then play the matrix game that their moves have set up. Compute the number of pure strategies for each player, and show that Player II has a strategy that prevents Player I from obtaining a positive gain.

9. Show that in Problem 8, Player I can be sure of winning at least 0 units.

10. Player II offers Player I a choice of  $N$  boxes,  $D$  of which contain a dollar bill each. Player I is to choose  $C$  of them, and his pay-off will be the number of dollars they contain. He is allowed to open any number of the boxes before making his choice, provided that he chooses only from among the boxes not yet opened. Player II knows the content of each box and may hand the boxes to Player I in any order. The pay-off to Player II is all the money not acquired by Player I. Determine an optimal strategy for each player.

11. Each of two players writes either 0, 1, or 2 on a piece of paper not visible to his opponent. Player I, knowing his own number, calls a guess of the sum of the two numbers. Next Player II, knowing his number, calls a guess of the sum, but he is not permitted to make the same guess as Player I. Player I receives one unit if his guess is correct; he pays one unit if Player II's guess is correct. If neither guess is correct, the game is a draw and the pay-off is zero. Set up the game and determine optimal strategies for each player.

12. A deck of cards is identified with the integers  $1, 2, \dots, N$ . After a thorough shuffling, Players I and II are each dealt a card. Player I may either keep his card or compel Player II to exchange cards with him. Player II may now either keep his card (which may be either the one he was dealt or the one he was compelled to accept), or exchange it for another taken out of the pack at random. When both players have made their choice, their hands are compared and the one with the higher hand receives one unit from his opponent. Solve the game for the case  $N = 5$ .

13. Each of two poker players places an ante of one unit and is dealt a single card from a standard deck. Player I, knowing what card he holds, has the option of folding immediately or betting any one of the amounts  $a_1, \dots, a_n$ . If he bets, then Player II has the option of folding (losing the ante) or seeing the bet. In this case the hands are compared, and the player with the higher hand wins. (Assume that the hands are identified

with the numbers  $1, \dots, 52$ , and that hand  $i$  is superior to hand  $j$  if  $i > j$ .) Solve the game when  $n = 1$  and  $a_1 > 1$ .

14. Solve Problem 13 for arbitrary  $n$ , assuming that  $1 < a_1 < \dots < a_n$ . (See also the general poker models of Volume II Chapter 9.)

15. In an attack-and-defense problem each player has two strategies, and the probability of success for the attack is given by a matrix  $\|a_{ij}\|$  with  $a_{11} > a_{21}$  and  $a_{12} < a_{22}$ . In addition, the attacker can (if he chooses) use part of his forces for reconnaissance. If this is done, his probability of success is decreased by a fixed amount  $r$ , but the defender's strategy is discovered unless he assigns part of his forces to counter-reconnaissance. This diversion of his forces increases the probability of a successful attack by an amount  $c$ . Set up the complete strategy spaces and the pay-off matrix for this game. Use dominance arguments to solve the game.

#### NOTES AND REFERENCES TO CHAPTER 4

**§ 4.1.** More elaborate Blotto games can be found in Blackett [27].

**§ 4.2.** The I.F.F. example and its ramifications (Problem 3), as described here, were proposed and solved by the members of the Rand Corporation; see Dresher [66].

**§ 4.3.** This poker model is due to Kuhn [156]. He did not, however, obtain a complete enumeration of all the optimal strategies of the game, as we do in this section. Kuhn's method was to set up the game initially in terms of the pay-off matrix corresponding to pure strategies and then to reduce the analysis, by dominance arguments, to consideration of a subclass of behavior strategies. Our method, by contrast, is based on a combination of intuition and direct analysis, a common way of dealing with general parlor games; see also Volume II Chapter 9.

**§ 4.4.** The election campaign model was proposed by Dresher [66].

**§ 4.5.** This bargaining example was developed at the Rand Corporation as a model of attack and defense; see Dresher [66].

**§ 4.6.** Problems 4 through 7 are various cases of the poker model of § 4.3. Problem 10 is a special case of a type of game discussed by Gale [95], involving repeated play of games with partial information structure. Problem 12 embodies the essential elements of the classical Le Her game, in which the standard card deck is replaced by a deck consisting of the number sequence  $1, 2, \dots, N$ . The original Le Her game was solved by Dresher [69].

## SOLUTIONS TO PROBLEMS OF CHAPTERS 1-4

## CHAPTER 1

*Problem 2.* A strategy for Player I consists of two components, the first giving the probability of betting if he draws Lo, and the second the probability of betting if he draws Hi. A strategy for Player II is given by a single component, the probability of calling if Player I bets. The optimal strategies are

$$\left\{ \frac{b-a}{b+a}, 1 \right\} \quad \text{and} \quad \left\{ \frac{2a}{b+a} \right\},$$

respectively. The value is  $a(b-a)/(b+a)$ .

*Problem 4.* Let  $\mathbf{A} = \|a_{ij}\|$  ( $i, j = 1, 2, \dots, 9$ ) denote the pay-off matrix. Then

$$a_{ij} = \begin{cases} -1 & \text{if } j - i \geq 2, \\ +1 & \text{if } i - j \geq 2, \\ +2 & \text{if } j - i = 1, \\ -2 & \text{if } i - j = 1, \\ 0 & \text{if } i - j = 0. \end{cases}$$

*Problem 6.* The optimal strategies are  $\langle \frac{5}{7}, \frac{2}{7} \rangle$  and  $\langle \frac{5}{7}, \frac{2}{7}, 0, 0 \rangle$ ; the value is  $\frac{13}{7}$ .

*Problem 8.* In both  $\mathbf{A}_1$  and  $\mathbf{A}_2$  the element 1 is a saddle point. Hence  $v(\mathbf{A}_1) = v(\mathbf{A}_2) = 1$  and  $v(\mathbf{A}_1) + v(\mathbf{A}_2) = 2$ . Setting  $x = 3$  yields  $v(\mathbf{A}_1 + \mathbf{A}_2) = 3 > 2$ , and setting  $x = -1$  yields  $v(\mathbf{A}_1 + \mathbf{A}_2) = 1 < 2$ .

*Problem 9.* Consider the game

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}.$$

The value is 1. However, note that  $\mathbf{x}^* = \langle \frac{1}{2}, \frac{1}{2} \rangle$ ,  $\mathbf{y}^* = \langle 1, 0 \rangle$  yields  $(\mathbf{x}^*, \mathbf{A}\mathbf{y}^*) = 1$  and yet  $\mathbf{y}^*$  is not optimal.

*Problem 10.* Let  $\mathbf{u} = \langle u_1, \dots, u_n \rangle$  be an optimal strategy for Player I and  $\mathbf{w} = \langle w_1, \dots, w_m \rangle$  be an optimal strategy for Player II. Then

$$\sum_{i=1}^n u_i a_{ij} \geq v \geq \sum_{j=1}^m a_{ij} w_j$$

for all  $i$  and  $j$ . It follows that if  $v \geq 0$  the first alternative holds, while if  $v < 0$  the second alternative holds.

*Problem 13.* In the game with matrix

$$\mathbf{C} = \begin{vmatrix} a & b \\ 0 & d \end{vmatrix},$$

the element 0 is a saddle point and  $\xi = 0$ ,  $\eta = 1$  are unique optimal strategies.

Thus

$$\mathbf{x} = \langle 0, 0, \dots, 0, x_{n_1+1}^*, \dots, x_{n_1+n_2}^* \rangle,$$

$$\mathbf{y} = \langle y_1^0, \dots, y_{m_1}^0, 0, \dots, 0 \rangle.$$

If Player II uses any of the first  $m_1$  columns, Player I obtains zero; if Player II uses any of the last  $m_2$  columns, Player I obtains at least  $d > 0$ . Thus, by taking combinations,  $C(\mathbf{x}, \mathbf{y}) \geq 0$  for any  $\mathbf{y}$ . Similarly  $C(\mathbf{x}, \mathbf{y}) \leq 0$  for any  $\mathbf{x}$ , and  $\{\mathbf{x}, \mathbf{y}\}$  is optimal.

*Problem 15.* Consider the game with matrix  $a_{ij} = f(\mathbf{x}_i, \mathbf{y}_j)$ . If  $\xi = \langle \xi_1, \dots, \xi_n \rangle$  and  $\eta = \langle \eta_1, \dots, \eta_m \rangle$  are optimal strategies, then

$$\mathbf{x}^0 = \sum_{i=1}^n \xi_i \mathbf{x}_i \quad \text{and} \quad \mathbf{y}^0 = \sum_{j=1}^m \eta_j \mathbf{y}_j$$

satisfy the inequalities.

*Problem 16.* If there exists no  $\mathbf{x}$  in  $X$  for which  $f(\mathbf{x}, \mathbf{y}) > \alpha$  for all  $\mathbf{y}$  in  $Y$ , then the set  $B_x = \{\mathbf{y} | f(\mathbf{x}, \mathbf{y}) \leq \alpha\}$  is nonvoid, closed because  $f$  is lower-semicontinuous, and convex because  $f$  is convex in  $\mathbf{y}$ . For any finite selection  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  the set  $B_{x_1, x_2, \dots, x_n} = \{\mathbf{y} | f(\mathbf{x}_i, \mathbf{y}) \leq \alpha \text{ all } i\}$  is also nonvoid, closed, and convex. Otherwise the image of  $Y$  in  $E^n$  defined by the mapping  $\mathbf{y} \rightarrow \{f(\mathbf{x}_i, \mathbf{y}) - \alpha\}$  has a convex closure which never touches the negative orthant. A strict separating plane exists with components of the normal  $\xi_i$  ( $i = 1, 2, \dots, n$ ) satisfying  $\xi_i \geq 0$  and not all zero. It follows that  $B_{x_0}$  is empty where

$$\mathbf{x}_0 = \sum_{i=1}^n \xi_i \mathbf{x}_i,$$

a contradiction. Hence the intersection  $\bigcap_{\mathbf{x}_i} B_{x_1, x_2, \dots, x_n}$  taken over all finite collections of  $\mathbf{x}_i$  contains an element  $\mathbf{y}_0$ , satisfying  $f(\mathbf{x}, \mathbf{y}_0) \leq \alpha$  for all  $\mathbf{x}$  in  $X$ .

*Problem 17.* Let  $\mathbf{A} = \|a_{ij}\|_{n \times m}$ ,  $b_j = \max_{1 \leq i \leq n} a_{ij}$ , and  $i_j$  be a value of  $i$  such that  $a_{i_j j} = b_j$ . We proceed by induction on  $m$ . If  $m = 2$ , consider the  $2 \times 2$  submatrix

$$\begin{vmatrix} a_{i_1 1} & a_{i_1 2} \\ a_{i_2 1} & a_{i_2 2} \end{vmatrix}.$$

(If  $i_1 = i_2$ , use this row and any other; if  $n = 1$ , the theorem is trivial.) If  $a_{i_1 1}$  is the saddle point of this  $2 \times 2$  matrix, then  $a_{i_1 1} \leq a_{i_1 2}$  and  $a_{i_1 1} = b_1 \geq a_{i_1 1}$  for all  $i$ , and hence  $a_{i_1 1}$  is a saddle point of the original  $n \times 2$  matrix. If  $a_{i_2 1}$  is the saddle point of the  $2 \times 2$  matrix, then  $a_{i_2 1} \leq a_{i_2 2}$  and  $a_{i_2 1} \geq a_{i_1 1} = b_1 \geq a_{i_2 1}$  for all  $i$ , and hence  $a_{i_2 1}$  is a saddle point of the original  $n \times 2$  matrix. Similar arguments hold for  $a_{i_1 2}$  and  $a_{i_2 2}$ .

Suppose now that the theorem is true for  $m - 1$ , and let  $a_{i_0 j_0}$  be a saddle point for the  $n \times (m - 1)$  submatrix consisting of the first  $m - 1$  columns of our  $n \times m$  matrix. If  $a_{i_0 j_0} \leq a_{i_0 m}$ , then  $a_{i_0 j_0}$  will be a saddle point of the full  $n \times m$  matrix. Hence we consider  $a_{i_0 j_0} > a_{i_0 m}$ . If  $a_{i_0 m} = a_{i_m m}$ , then  $a_{i_0 m}$  is a

saddle point of the  $n \times m$  matrix. Hence we may also consider only the case  $a_{i_0m} < a_{i_mm}$ .

*Case I:*  $a_{i_mm} \leq a_{i_mj_0}$ . Let  $k$  be an arbitrary integer between 1 and  $n - 1$ . Consider the  $2 \times 2$  submatrix

$$\begin{vmatrix} a_{i_mk} & a_{i_mm} \\ a_{i_0k} & a_{i_0m} \end{vmatrix}.$$

Now  $a_{i_0m} < a_{i_0j_0} \leq a_{i_0k}$ , and thus  $a_{i_0k}$  cannot be a saddle point of this  $2 \times 2$  submatrix. Also since  $a_{i_0m} < a_{i_mm}$ , it follows that  $a_{i_0m}$  cannot be a saddle point. If  $a_{i_mk}$  is a saddle point, then  $a_{i_mk} \geq a_{i_0k} \geq a_{i_0j_0} \geq a_{i_mj_0} \geq a_{i_mm}$ . On the other hand, if  $a_{i_mm}$  is a saddle point, we also have  $a_{i_mk} \geq a_{i_mm}$ . Thus  $a_{i_mm}$  is a saddle point of the  $n \times m$  matrix.

*Case II:*  $a_{i_mm} > a_{i_mj_0}$ . We proceed as above; in case  $a_{i_mk}$  is the saddle point of the  $2 \times 2$  submatrix, we have  $a_{i_mk} \geq a_{i_mj_0}$  as above. If  $a_{i_mm}$  is the saddle point, then  $a_{i_mk} \geq a_{i_mm} > a_{i_mj_0}$ . Thus  $a_{i_mj_0} \leq a_{i_mk}$  ( $k = 1, 2, \dots, m$ ). Now consider the  $2 \times 2$  submatrix

$$\begin{vmatrix} a_{i_mj_0} & a_{i_mm} \\ a_{i_0j_0} & a_{i_0m} \end{vmatrix}.$$

Clearly  $a_{i_0m} < a_{i_0j_0}$ , hence  $a_{i_0j_0}$  is not a saddle point of this  $2 \times 2$ ;  $a_{i_0m} < a_{i_mm}$ , hence  $a_{i_0m}$  is not a saddle point; and  $a_{i_mm} > a_{i_mj_0}$ , hence  $a_{i_mm}$  is not a saddle point. Thus  $a_{i_mj_0}$  must be the saddle point of this  $2 \times 2$  submatrix, and hence  $a_{i_mj_0} \geq a_{i_0j_0} \geq a_{ij_0}$  ( $i = 1, 2, \dots, n$ ). Thus  $a_{i_mj_0}$  is a saddle point of the  $n \times m$  matrix.

*Problem 18.* The example

$$\begin{vmatrix} 3 & 0 & 4 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{vmatrix}$$

shows that the result is not correct if the strictness assumption is removed.

*Problem 19.* See Wolfe [255].

## CHAPTER 2

*Problem 3.* The optimal strategy for each player is unique:

$$\mathbf{x}^* = \mathbf{y}^* = \left\langle \frac{1/a_1}{\sum_i 1/a_i}, \frac{1/a_2}{\sum_i 1/a_i}, \dots, \frac{1/a_n}{\sum_i 1/a_i} \right\rangle; v = \frac{1}{\sum_i 1/a_i}.$$

*Problem 8.*  $\langle 0, 0, 0, 0, \frac{1}{16}, \frac{5}{16}, \frac{4}{16}, \frac{5}{16}, \frac{1}{16} \rangle$ .

*Problem 10.*  $v = \frac{10}{3}$ ;  $\mathbf{x}^0 = \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$  is the unique optimal strategy for Player I. Extreme-point optimal strategies for Player II are:

$$\begin{aligned} & \langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0 \rangle; \langle \frac{1}{2}, \frac{1}{3}, 0, \frac{1}{6}, 0, 0 \rangle; \langle \frac{4}{9}, \frac{4}{9}, 0, 0, \frac{1}{9}, 0 \rangle; \\ & \langle \frac{1}{3}, 0, 0, \frac{1}{2}, 0, \frac{1}{6} \rangle; \langle 0, \frac{4}{15}, 0, 0, \frac{7}{15}, \frac{4}{15} \rangle; \langle 0, \frac{1}{9}, \frac{7}{9}, 0, 0, \frac{1}{9} \rangle; \\ & \langle 0, 0, \frac{2}{3}, \frac{1}{6}, 0, \frac{1}{6} \rangle; \langle 0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle. \end{aligned}$$

*Problem 13.* Three cases must be distinguished:  $n = 3k$ ,  $n = 3k + 1$ , and  $n = 3k + 2$ . If  $n = 3k$ ,

$$\mathbf{x}^0 = \left\langle 0, \frac{1}{k}, 0, 0, \frac{1}{k}, 0, \dots, 0, \frac{1}{k}, 0 \right\rangle,$$

$$\mathbf{y}^0 = \left\langle \frac{1}{k}, 0, 0, \frac{1}{k}, 0, 0, \frac{1}{k}, 0, \dots, \frac{1}{k}, 0, 0 \right\rangle,$$

$$v = \frac{1}{k}.$$

The other cases are similar. The optimal strategies are not unique.

*Problem 16.* Column 2 dominates column 3; after deleting column 2, row 1 dominates row 3. It follows that

$$v = \frac{a_4a_5 - a_1a_3}{a_4 + a_5 - a_1 - a_3};$$

$$\mathbf{x}^0 = \left\langle \frac{a_5 - a_1}{a_4 + a_5 - a_1 - a_3}, \frac{a_4 - a_3}{a_4 + a_5 - a_1 - a_3}, 0 \right\rangle;$$

$$\mathbf{y}^0 = \left\langle \frac{a_5 - a_3}{a_4 + a_5 - a_1 - a_3}, 0, \frac{a_4 - a_1}{a_4 + a_5 - a_1 - a_3} \right\rangle.$$

In the original matrix,  $A(\mathbf{x}^0, \beta_2) > v$ ; hence  $y_2 = 0$ . Similarly,  $x_3 = 0$ . Since  $v = A(\alpha_1, \mathbf{y}^0) = a_4y_1^0 + a_3y_3^0$ , and since  $y_1^0 > 0$ ,  $y_3^0 > 0$ , and  $y_1^0 + y_3^0 = 1$ , it follows that  $a_3 < v < a_4$ . By virtue of Lemma 2.1.2,  $a_4x_1^0 + a_1x_2^0 = v$  and  $a_3x_1^0 + a_5x_2^0 = v$ , and similarly for  $\mathbf{y}^0$ . These relations uniquely determine  $\mathbf{x}^0$  and  $\mathbf{y}^0$ . Now clearly we have  $x_1^0 > y_1^0 > x_2^0$ , and from  $x_1^0 > x_2^0$  and  $x_1^0 + x_2^0 = 1$  it follows that  $x_1^0 > \frac{1}{2}$ .

*Problem 18.* Let  $\mathbf{x}^0 \in X^k$ . Then  $G^k(\mathbf{x}^0, \mathbf{y}) \geq v^k$  for all  $\mathbf{y}$  ( $k = 1, \dots, K$ ). Now if Player II uses  $\mathbf{y}$  in the game  $G$ , and Player I uses  $\mathbf{x}^0$ ,

$$\begin{aligned} G(\mathbf{x}^0, \mathbf{y}) &= \sum_{i,j} x_i^0 G_{ij} y_j = \sum_{i,j} x_i^0 \sum_k G_{ij}^k y_j = \sum_k \sum_{i,j} x_i^0 G_{ij}^k y_j \\ &= \sum_k G^k(\mathbf{x}^0, \mathbf{y}) \geq \sum_k v^k. \end{aligned}$$

Hence

$$v \geq \sum_k v^k.$$

*Problem 20.* Suppose that  $\mathbf{A}$  is c.m. symmetric of even order. By Theorem 2.5.1, the kernel must be the entire matrix. Since the value of a symmetric game is zero,  $|\mathbf{A}| = 0$  (Theorem 2.4.3). Since  $\mathbf{A}$  is singular and skew-symmetric of even order, it has rank at most  $n - 2$  (since the rank must be even). But the fact that  $\mathbf{A}$  is c.m. implies that  $\mathbf{A}\mathbf{y} = 0$  has a solution with all components strictly positive. Since the rank is at most  $n - 2$ , we may perturb the solution slightly and still have a positive solution with components adding up to 1, thus contradicting the uniqueness.

*Problem 22.* Proof by induction on the number of columns.

## CHAPTER 3

*Problem 3.* By (3.3.3)  $\dim \overline{X} = r(\tilde{\mathbf{A}}) + \dim \eta(\tilde{\mathbf{A}}') = 1$ . From Theorem 3.3.1,  $\dim X^0 + 1 = \dim \eta(\tilde{\mathbf{A}}')$ . Combining these results yields  $\dim \overline{X} - \dim X^0 = r(\tilde{\mathbf{A}})$ . By Problem 2,  $\tilde{\mathbf{A}}$  is skew-symmetric and hence has even rank.

*Problem 4.* See [98].

*Problem 6.*

$$\begin{vmatrix} 2 & 1 & 2 & -1 & 1 \\ -2 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 & -2 \end{vmatrix}$$

*Problem 8.* Let  $\mathbf{y}$  and  $\epsilon$  be given; then choose  $N$  so that  $y_1 + \dots + y_N > 1 - \epsilon$ . Now choose  $i$  so that  $i > N/\epsilon$ , and let  $\mathbf{x}$  assign mass 1 to  $i$ . Then  $A(\mathbf{x}, \mathbf{y}) = \sum_j a_{ij}y_j \geq (1 - \epsilon)a_{iN} + \epsilon(-1)$ . But  $a_{iN} \geq 1 - \epsilon$ ; hence  $A(\mathbf{x}, \mathbf{y}) > 1 - 3\epsilon$ , and of course  $A(\mathbf{x}, \mathbf{y}) \leq 1$ . Thus

$$\inf_y \sup_x A(\mathbf{x}, \mathbf{y}) = 1.$$

Similarly,

$$\sup_x \inf_y A(\mathbf{x}, \mathbf{y}) = -1.$$

*Problem 10.* Use the graphical representation procedure of Section 2.3.

*Problem 11.* Let  $\{\mathbf{x}^0, \mathbf{y}^0\}$  be optimal for  $\mathbf{A}$  and  $\{\mathbf{x}_\alpha, \mathbf{y}_\alpha\}$  be optimal for  $\mathbf{A} + \alpha\mathbf{H}$ . Then

$$v(\mathbf{A} + \alpha\mathbf{H}) \leq (\mathbf{x}_\alpha, (\mathbf{A} + \alpha\mathbf{H})\mathbf{y}^0) \leq v(\mathbf{A}) + \alpha(\mathbf{x}_\alpha, \mathbf{H}\mathbf{y}^0)$$

and

$$v(\mathbf{A} + \alpha\mathbf{H}) \geq (\mathbf{x}^0, (\mathbf{A} + \alpha\mathbf{H})\mathbf{y}_\alpha) \geq v(\mathbf{A}) + \alpha(\mathbf{x}^0, \mathbf{H}\mathbf{y}_\alpha).$$

Hence

$$(\mathbf{x}^0, \mathbf{H}\mathbf{y}_\alpha) \leq \frac{v(\mathbf{A} + \alpha\mathbf{H}) - v(\mathbf{A})}{\alpha} \leq (\mathbf{x}_\alpha, \mathbf{H}\mathbf{y}^0).$$

Now let  $X^*$  be the set of all possible limit points of sequences  $\{\mathbf{x}_{\alpha_k}\}$  as  $\alpha_k \rightarrow 0$ , and let  $Y^*$  be defined similarly with respect to the  $\mathbf{y}_\alpha$ . By Lemma 3.5.1,  $X^* \subset X^0(\mathbf{A})$  and  $Y^* \subset Y^0(\mathbf{A})$ . We clearly have

$$\begin{aligned} \max_{\mathbf{x} \in X^0(\mathbf{A})} \min_{\mathbf{y} \in Y^0(\mathbf{A})} (\mathbf{x}, \mathbf{H}\mathbf{y}) &\leq \max_{\mathbf{x} \in X^0(\mathbf{A})} \min_{\mathbf{y} \in Y^*} (\mathbf{x}, \mathbf{H}\mathbf{y}) \leq \lim_{\alpha \rightarrow 0^+} \frac{v(\mathbf{A} + \alpha\mathbf{H}) - v(\mathbf{A})}{\alpha} \\ &\leq \max_{\mathbf{x} \in X^*} \min_{\mathbf{y} \in Y^0(\mathbf{A})} (\mathbf{x}, \mathbf{H}\mathbf{y}) \leq \max_{\mathbf{x} \in X^0(\mathbf{A})} \min_{\mathbf{y} \in Y^0(\mathbf{A})} (\mathbf{x}, \mathbf{H}\mathbf{y}). \end{aligned}$$

*Problem 16.* First we may confine our attention to the subset of the space of matrices where no two elements are equal. This subset has probability 1. If no two elements of the matrix are equal, there can be only one saddle point. Hence the events  $\{a_{ij} \text{ is a saddle point}\}$  are disjoint events. Thus

$$\Pr \{ \mathbf{A} \text{ has a saddle point} \} = \sum_{i,j} \Pr \{ a_{ij} \text{ is a saddle point} \}$$

(Pr denotes probability of the event indicated). But any  $a_{ij}$  is a saddle point if it is the largest element in its column and the smallest in its row, hence if it is the  $m$ th ranking in a collection of  $m + n - 1$  arbitrary choices of numbers. The probability of this is

$$\frac{(n-1)!(m-1)!}{(m+n-1)!}.$$

Hence

$$\Pr \{ \mathbf{A} \text{ has a saddle point} \} = m \cdot n \frac{(m-1)!(n-1)!}{(m+n-1)!} = \frac{m!n!}{(m+n-1)!}.$$

*Problem 17.* See Thompson [231].

*Problem 19.* Suppose that there is no optimal  $\mathbf{x}^0$  which uses at most  $n - 1$  indices. Then all optimal  $\mathbf{x}^0$  use all  $n$  indices. But any  $\mathbf{x}^0$  using all  $n$  indices must have the kernel corresponding to all of  $\mathbf{A}$ . Hence  $\dim X^0 = 0$  and  $\dim \bar{X} = n - 1$ . It follows that  $\mathbf{y}^0$  is unique, and hence that  $\dim Y^0 = 0$ . Consequently, by Theorem 3.1.2,  $\dim \bar{Y} = n - 1$ . But this means that  $\{\mathbf{y}^0\} = Y^0$  is interior to  $Y$ , a contradiction.

*Problem 20.* Use Theorem 3.5.1. (See also Lemma 5.3.1)

#### CHAPTER 4

*Problem 2.*  $\mathbf{x}^0 = \mathbf{y}^0 = \langle 1/n, \dots, 1/n \rangle; v = (n-2)/n$ .

*Problem 3.* There are  $2^n$  possible signals. The only reasonable strategies for the observer are (1) to identify all signalers as enemy planes; (2) to identify planes giving the agreed-upon signal as friendly and all others as enemy planes. If (1) is used, the expected pay-off is  $pb + (1-p)c$  regardless of the agreed-upon signal. If (2) is adopted, the pay-off is  $pa + (1-p)d$  if the enemy uses the same signal, and  $pa + (1-p)c$  if the enemy uses any other signal. Hence the matrix

		enemy signals		
		1	...	$2^n$
Plan (2)	Code 1	$pa + (1-p)d \dots pa + (1-p)c$		
	2	$pa + (1-p)c \dots pa + (1-p)c$		
	$\vdots$	$\vdots$		$\vdots$
	$2^n$	$pa + (1-p)c \dots pa + (1-p)d$		
Plan (1): any code		$pb + (1-p)c \dots pb + (1-p)c$		

Note that if

$$pa + (1-p) \left( \frac{d + (2^n - 1)c}{2^n} \right) > pb + (1-p)c, \quad (\text{a})$$

the last row will be dominated; hence (by symmetry)  $\langle 1/2^n, \dots, 1/2^n \rangle$  is the unique optimal strategy for each player. On the other hand, if the opposite inequality holds, Player I can assure himself of the amount  $pb + (1 - p)c$  by using plan (1). If Player I uses any other row, Player II can hold him to the amount of the right side of Eq. (a) by using  $\langle 1/2^n, \dots, 1/2^n \rangle$ . Hence, in this case, the observer will identify all signals as enemy signals and the enemy can use any strategy whatsoever. The inequality (a) reduces to that of the problem.

*Problem 7.*

$$\begin{array}{ll}
 x_3^0 = 1 & y_3^0 = 1 \\
 x_2^0 = 0 & y_2^0 = 0 \\
 x_1^0 = 0 & y_1^0 = 0 \\
 \alpha_2^0 = 0 & \beta_2^0 = 0 & \gamma_2^0 = 1 & \alpha_3^0 + \beta_3^0 + \gamma_3^0 = 1 \\
 \alpha_1^0 = 0 & \beta_1^0 = 0 & \gamma_1^0 = 1 & -2a\gamma_3^0 + b\beta_3^0 \geq 2a
 \end{array}$$

*Problem 9.* One such strategy for Player I is as follows: Place 3 in the square  $(1, 1)$ . If II doesn't play in  $(1, 2)$ , place 2 in  $(1, 2)$  and in the resulting game play  $\langle 1, 0, 0 \rangle$ . If II plays in  $(1, 2)$ , place 2 in  $(3, 2)$ . If II doesn't play in  $(3, 3)$  at his second move, place 1 in  $(3, 3)$  and in the resulting game play  $\langle 0, 0, 1 \rangle$ . If II does play in  $(3, 3)$  at his second move, then place 1 in  $(2, 3)$ . The resulting matrix is

3	$-j_1$	0
$-j_3$	0	1
0	2	$-j_2$

The following strategy (for Player I) now yields zero:

$$\left\langle \frac{j_2 j_3}{3j_2 + 3 + j_2 j_3}, \frac{3j_2}{3j_2 + 3 + j_2 j_3}, \frac{3}{3j_2 + 3 + j_2 j_3} \right\rangle.$$

*Problem 11.* Optimal strategy for Player I: Pick 0, 1, and 2 each with probability  $\frac{1}{3}$  and guess 2.

Optimal strategy for Player II: Pick 0, 1, and 2 each with probability  $\frac{1}{3}$ . If 0 is chosen and I guesses 2, guess 0. If 1 or 2 is chosen and I guesses 2, guess 3. If I doesn't guess 2, guess 2.

*Problem 13.* Let  $\alpha_i$  = probability I bets on card  $i$  and  $\beta_j$  = probability II sees on card  $j$ . The pay-off matrix becomes

$$K(\alpha, \beta) = \frac{1}{51.52} \sum_{i \neq j} \{ -(1 - \alpha_i) + \alpha_i(1 - \beta_j) + \alpha_i \beta_j L(i, j) a \}.$$

The solution is obtained by employing reasoning analogous to that employed in

Section 4.3. Any optimal strategy has the form

$$\beta_1^0 = \dots = \beta_{m-1}^0 = 0; \beta_m^0 = m - 52 + \frac{2(51)}{a+1}; \beta_{m+1}^0 = \dots = \beta_{52}^0 = 1$$

and

$$\alpha_m^0 = \dots = \alpha_{52}^0 = 1 \quad \text{with} \quad \alpha_1^0, \dots, \alpha_{m-1}^0 \quad \text{arbitrary}$$

between zero and one except satisfying

$$\sum_{i=1}^{m-1} \alpha_i^0 = \frac{a-1}{a+1} (52 - m) \quad \text{if } \beta_m^0 < 1$$

and satisfying

$$\sum_{i=1}^{m-1} \alpha_i^0 \geq \frac{a-1}{a+1} (52 - m) \quad \text{and} \quad \sum_{i=1}^{m-2} \alpha_i^0 \leq \frac{a-1}{a+1} (53 - m) \quad \text{if } \beta_m^0 = 1.$$

$m$  is equal to the largest integer not exceeding  $53 - 2 \cdot 51/(a+1)$ .

*Problem 15.* In addition to the initial strategies labeled 1 and 2, the defender now has two new strategies that may be denoted by 1D and 2D. Here 1D represents the strategy where the defender uses his defense 1 against attack and takes counter-measures against reconnaissance. The attacking player, in addition to his initial strategies labeled 1 and 2 has available strategies where he reconnoiters and utilizes the information obtained; for example 1R may denote the strategy 1 unless he knows that the enemy is using strategy 2. Similarly, we have the strategy 2R. There are other possible strategies for player I that ignore or misuse the information obtained by reconnaissance. These are all dominated. The relevant pay-off matrix reduces to

	1	2	1D	2D
1	$a_{11}$	$a_{12}$	$a_{11} + c$	$a_{12} + c$
2	$a_{21}$	$a_{22}$	$a_{21} + c$	$a_{22} + c$
1R	$a_{11} - r$	$a_{22} - r$	$a_{11} - r + c$	$a_{12} - r + c$
2R	$a_{11} - r$	$a_{22} - r$	$a_{21} - r + c$	$a_{22} - r + c$

## CHAPTER 5

### LINEAR PROGRAMMING

Programming is concerned with determining feasible programs (plans, schedules, allocations) that are optimal with respect to an agreed-upon criterion. Determining an optimal program is usually a matter of selecting from among all feasible programs a program that maximizes (or minimizes) the criterion function.

A feasible program is one that satisfies certain constraints. When these constraints are linear inequalities and the objective (criterion) function is also a linear expression, we have a problem in linear programming. When the objective function to be optimized or the defining constraint relations are nonlinear, we have a problem in nonlinear programming.

The assumptions of linearity, finiteness of the number of variables, and additivity of the gains and losses limit the applications of linear programming. The merits and demerits of these assumptions are not discussed here. In this chapter we are concerned only with the nature of linear programming problems and the properties their solutions must possess.

It might seem that an immediate solution to our problem would be furnished by the calculus, with its standard methods of determining the maxima and minima of functions. Unfortunately, the ordinary calculus can rarely be applied to programming problems, since their solutions fall on the boundaries of the variables domain. Hence we need new techniques for analyzing the qualitative structure of such problems.

Moreover, because of the vast number of practical applications of this theory, it is important to find efficient computational procedures for approximating solutions. This problem will be fully discussed in Chapter 6. The present chapter emphasizes the simple elegance of the theory.

**5.1 Formulation of the linear programming problem.** The linear programming problem has the following mathematical form:

Given that the variables  $x_1, \dots, x_n$  satisfy the inequalities

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m, \\ x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0, \end{aligned} \tag{5.1.1}$$

find the set of values of  $x_1, \dots, x_n$  which maximizes\*

$$(\mathbf{c}, \mathbf{x}) = x_1c_1 + x_2c_2 + \dots + x_nc_n.$$

Expressed in vector notation, the problem is to find that  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  which maximizes  $(\mathbf{c}, \mathbf{x})$  subject to the constraints  $\mathbf{x} \geq \mathbf{0}$  and  $\mathbf{Ax} \leq \mathbf{b}$ , where  $\mathbf{c} = \langle c_1, \dots, c_n \rangle$ ,  $\mathbf{b} = \langle b_1, \dots, b_m \rangle$ , and  $\mathbf{A}$  is the matrix of coefficients of the system of linear inequalities (5.1.1).

A vector  $\mathbf{x}$  which satisfies the constraints (5.1.1) is said to be a feasible vector (strategy). The set of all feasible vectors is a polyhedral convex set, since it is the intersection of the  $m + n$  half-spaces described by (5.1.1). It may happen that the inequalities are inconsistent, so that there is no feasible vector. Since the contour lines of the function  $(\mathbf{c}, \mathbf{x})$  are hyperplanes in  $n$ -space, the maximum point of  $(\mathbf{c}, \mathbf{x})$  cannot be an interior point of one of the bounding surfaces. Consequently, the maximum is achieved either at a single vertex or at all points on some bounding face. It may happen, of course, that even when the feasible set is nonvoid the problem has no solution, since  $(\mathbf{c}, \mathbf{x})$  could become infinite as  $\mathbf{x}$  tends to the boundary of the feasible set. The reader may readily construct such examples.

Many important problems in economics and management science can be expressed in the linear programming form without significant distortion or loss of generality. The remainder of this section will be devoted to outlining several such problems.

A. *Diet problem.* There are  $m$  different types of food ( $i = 1, 2, \dots, m$ ) that supply varying quantities of the  $n$  nutrients ( $j = 1, 2, \dots, n$ ) that are essential to good health. The  $n$  nutrients may be thought of as protein, minerals, vitamins, etc.

Let  $y_i$  be the number of units of food  $i$  bought, and let  $b_i$  be the price per unit of food  $i$ . Let  $c_j$  be the minimum requirement of nutrient  $j$ .

Let  $a_{ij}$  be the amount of nutrient content  $j$  in one unit of food  $i$ .

The question is, how can the requisite nutrients be supplied at minimum cost? That is, subject to the constraint

$$\sum_{i=1}^m y_i a_{ij} \geq c_j \quad (j = 1, 2, \dots, n),$$

we seek to determine a vector  $\mathbf{y}$  which minimizes the total cost

$$\sum_{i=1}^m b_i y_i.$$

---

\* Or minimizes. If the problem is one of minimization, the vector inequality  $\mathbf{Ax} \leq \mathbf{b}$  obtained from (5.1.1) is usually reversed and takes the form  $\mathbf{yA} \geq \mathbf{c}$ . In this case, the objective function is changed from  $(\mathbf{c}, \mathbf{x})$  into  $(\mathbf{y}, \mathbf{b})$ . See also the first paragraph of Section 5.2.

Historically, this is one of the earliest linear programs considered. The problem was constructed and solved on the basis of prices for the year 1944, and it was found that approximately \$60 a year would finance an adequate diet—adequate, at least, for such rare souls as might take pleasure in an endless round of dried navy beans, cabbage, wheat flour, and the like. Despite this inauspicious beginning, the trial-and-error method used to solve the diet problem proved to be the forerunner of a number of successful linear programming analyses of industrial and management problems. The original analysis is currently used in deciding on the best feed for chickens.

Linear programming problems which closely resemble the diet problem occur in the oil industry—for example, in calculating the cheapest way to blend available quantities and qualities of gasoline into a required assortment of fuels having specified minimum qualities.

B. *Transportation problem.* There are  $m$  ports or centers of supply of a certain commodity, and  $n$  destinations or markets to which this commodity must be shipped. The  $i$ th port possesses an amount  $s_i$  of the commodity ( $i = 1, 2, \dots, m$ ), and the requirements are such that the  $j$ th destination is to receive the amount  $r_j$  ( $j = 1, 2, \dots, n$ ).

Let  $x_{ij}$  be the quantity of the commodity we intend to ship from port  $i$  to destination  $j$ .

Let  $c_{ij}$  be the cost of shipping one unit from port  $i$  to destination  $j$ .

Then, subject to  $\sum_j x_{ij} \leq s_i$  and  $\sum_i x_{ij} \geq r_j$ , we seek to minimize the total shipping costs

$$\sum_{i,j} x_{ij} c_{ij}.$$

This transportation model has received more attention than any other linear programming problem; many special computational schemes have been designed exclusively for it. Several linear programming examples arising from apparently different backgrounds—notably the optimal assignment problem and the railway flow problem—are reducible to the transportation problem. The model is investigated in detail in Section 5.9.

C. *Warehouse storage problem.* Suppose that  $C$  is the capacity of a warehouse, and let  $A$  represent the initial stock size. Now let  $x_i$  represent the amount of the commodity bought at the end of the  $i$ th period at a unit price of  $c_i$  dollars, and  $y_i$  represent the amount sold in the  $i$ th period at a unit price of  $b_i$  dollars. Finally, let  $n$  denote the total horizon (number of periods to be considered).

The problem is to determine the  $\mathbf{x}$  and  $\mathbf{y}$  that maximize the profit function

$$\sum_{i=1}^n (-c_i x_i + b_i y_i),$$

subject to the restrictions

$$y_{i+1} \leq A + \sum_{j=1}^i (x_j - y_j) \leq C$$

and

$$x_i \geq 0, y_i \geq 0 \quad (i = 1, 2, \dots, n).$$

The right-hand inequality in the first expression takes care of the capacity limitations on any buying and selling program. The left-hand inequality states that it is not possible to sell more than is available.

**D. Activity analysis problem.** There are  $n$  ways (activities) of producing an output from available stocks of  $m$  primary resources.

Let  $a_{ij}$  be the amount of the  $i$ th material used in operating the  $j$ th activity at unit intensity. The intensity of an activity can be related to the number of factories engaged in this one activity, or to the level of labor input, which often is considered as an exogenous resource; any such interpretation will serve.

Let  $b_i$  be the available stock of the  $i$ th resource.

Let  $c_j$  be the value of the output achieved by operating the  $j$ th activity at unit intensity.

Let  $x_j$  be the level of intensity of the  $j$ th activity to be undertaken.

Subject to the restrictions

$$\sum_{j=1}^n a_{ij}x_j \leq b_i$$

and  $x_j \geq 0$  for all  $i$  and  $j$ , respectively, we want to choose  $\mathbf{x}$  so as to maximize the total value of the output,

$$\sum_{j=1}^n c_jx_j.$$

**5.2 The linear programming problem and its dual.** In the previous section the linear programming problem was defined in terms of finding a vector  $\mathbf{x}$  such that  $\mathbf{Ax} \leq \mathbf{b}$ ,  $\mathbf{x} \geq \mathbf{0}$ , and  $(\mathbf{c}, \mathbf{x}) = \sum x_j c_j$  is maximized. This will be referred to as Problem I. A closely related problem, called the dual of Problem I, is to find a vector  $\mathbf{y} = \langle y_1, \dots, y_m \rangle$  that minimizes  $(\mathbf{y}, \mathbf{b}) = \sum b_i y_i$  and satisfies the constraints  $\mathbf{yA} \geq \mathbf{c}$  and  $\mathbf{y} \geq \mathbf{0}$ . ( $\mathbf{A}$  is an  $m \times n$  matrix,  $\mathbf{c} = \langle c_1, \dots, c_n \rangle$ ,  $\mathbf{b} = \langle b_1, \dots, b_m \rangle$ .)

In this dual problem, which we shall call Problem II, minimization has replaced maximization,  $\mathbf{b}$  and  $\mathbf{c}$  have been interchanged, and the inequality sign in the major constraints has been reversed. Formally, Problems I and II can be characterized as follows:

$$\begin{array}{ll}
 \text{Problem I: } \mathbf{x} = \langle x_1, \dots, x_n \rangle, & \text{Problem II: } \mathbf{y} = \langle y_1, \dots, y_m \rangle, \\
 \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}, & \mathbf{yA} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0}, \\
 \max_{\mathbf{x}} (\mathbf{c}, \mathbf{x}). & \min_{\mathbf{y}} (\mathbf{y}, \mathbf{b}).
 \end{array}$$

If some of the constraining relations in Problem I are equalities, then in Problem II the corresponding components of  $\mathbf{y}$  are allowed to vary unrestrictedly (i.e., they may be negative). Moreover, if in Problem I a certain subset of variables  $x_i$  vary without restriction, then the corresponding constraining relations of  $\mathbf{yA} \geq \mathbf{c}$  are equalities.

In linear programming problems involving economic ideas, the dual problem will usually possess an economic interpretation. In the activity analysis model, for example, the dual variables  $y_i$  are usually assigned the meaning of prices; it is common to refer to  $\mathbf{y}$  as the "shadow-price" vector, where shadow price may be interpreted as a fictitious price that society gives to a commodity when the economy is operating in a state of perfect competition. (See Chapter 8.) To illustrate these concepts, we return briefly to the activity analysis model and the transportation model of the previous section.

*Activity analysis model.* Let  $\mathbf{y}$  be the external shadow prices that a company should set on its resources in order to reflect their value to society. From the viewpoint of the whole economy, the company's resources would include (besides raw materials and labor) services of management, clerical services, etc. As previously, let  $c_j$  be the dollar value of the output achieved by operating activity  $j$  at unit intensity. Then, if  $y_i$  is the internal price of one unit of the  $i$ th resource,

$$\sum_i y_i a_{ij}$$

is the total dollar value (according to these prices) of the output resulting from operating  $j$  at unit intensity.

Hence the constraints  $\mathbf{yA} \geq \mathbf{c}$  state that internal prices cannot be set to get more value from a product than you put into it. In other words, in a situation of equilibrium the laws of economics for society require no profits. The nature of this assumption and its consequences are fully explored in our general discussion of competitive equilibrium in Chapter 8.

In the dual problem, subject to the constraints  $\mathbf{yA} \geq \mathbf{c}$  and  $\mathbf{y} \geq \mathbf{0}$ , we seek to minimize the value of the total resources,  $\sum_i b_i y_i$ . As previously,  $b_i$  represents the total amount of the  $i$ th resource in stock.

*Transportation model.* The dual constraints of the transportation model are of the form

$$\begin{array}{ll}
 c_{ij} \geq v_j - u_i, \\
 u_i \geq 0 & (i = 1, \dots, m), \\
 v_j \geq 0 & (j = 1, \dots, n).
 \end{array} \tag{5.2.1}$$

Here  $u_i$  can be interpreted as the shadow price (dollar or value) associated with sending one unit of commodity from port  $i$ , and  $v_j$  as the shadow price associated with receiving one unit of commodity at destination  $j$ . The meaning of equality in (5.2.1) is that under conditions of equilibrium the value to society of one unit of commodity received at destination  $j$  is absorbed in the shipping cost  $c_{ij}$  and the shadow cost of preparing one unit of commodity for shipment at port  $i$ . We shall see later that if the optimal policy calls for a shipment from  $i$  to  $j$ , then equality must hold in (5.2.1) corresponding to these indices.

The dual to the transportation problem is as follows: Subject to the restrictions of (5.2.1), we wish to select  $u_i$  and  $v_j$  which maximize

$$\sum_{i=1}^m u_i s_i - \sum_{j=1}^n v_j r_j,$$

where  $s_i$  is the amount available at source  $i$  and  $r_j$  is the amount demanded at destination  $j$ .

The relationships between the linear programming problem and its dual are developed in the following sections. (See also Problem 18, which embodies another meaning of the shadow price.)

**5.3 The principal theorems of linear programming (preliminary results).** This and the next two sections are primarily concerned with the proof of the two fundamental theorems of linear programming. In this section, we develop some preliminary results concerning convex sets and their application to the linear programming problem. Section 5.4 gives the conditions under which a solution to the problem exists and the relation between the solutions of Problems I and II. Section 5.5 explains the connection between linear programming and game theory.

In seeking a solution to a maximization problem subject to constraints, it is customary to examine the associated Lagrangian form. However, when we have inequality constraints instead of the equalities which are common in the study of variational problems, the analysis of the corresponding Lagrangian form becomes more difficult. Fortunately, the theory of convex sets is exactly what is needed to overcome the difficulties presented by linear inequality constraints. Accordingly, the succeeding Lemma 5.3.1 plays a vital role in the analysis of the Lagrangian form of Problem I. Using the notation of Section 5.2, we may define the Lagrangian form of this problem by

$$\phi(\mathbf{x}, \mathbf{y}) = (\mathbf{c}, \mathbf{x}) + (\mathbf{y}, \mathbf{b} - \mathbf{Ax}) \quad (5.3.1)$$

for  $\mathbf{x} = \langle x_1, \dots, x_n \rangle \geq 0$  and  $\mathbf{y} = \langle y_1, \dots, y_m \rangle \geq 0$ .

Problem II is converted into a form of Problem I if the signs of all the parameters are changed, i.e., if  $\mathbf{A}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  are replaced by  $-\mathbf{A}$ ,  $-\mathbf{b}$ , and  $-\mathbf{c}$ . Hence, the Lagrangian form associated with Problem II, consistent with (5.3.1), is defined as

$$\psi(\mathbf{y}, \mathbf{x}) = -(\mathbf{y}, \mathbf{b}) + (\mathbf{y}\mathbf{A} - \mathbf{c}, \mathbf{x}) \quad (5.3.2)$$

for  $\mathbf{x} \geq \mathbf{0}$  and  $\mathbf{y} \geq \mathbf{0}$ , which by rearrangement is seen to be  $-\phi(\mathbf{x}, \mathbf{y})$ .

A pair of vectors  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{y}}$  will be called a *saddle point* of  $\phi(\mathbf{x}, \mathbf{y})$  if

$$\bar{\mathbf{x}} \geq \mathbf{0}, \quad \bar{\mathbf{y}} \geq \mathbf{0}, \quad (5.3.3)$$

and

$$\phi(\mathbf{x}, \bar{\mathbf{y}}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \leq \phi(\bar{\mathbf{x}}, \mathbf{y}) \quad \text{for all } \mathbf{x} \geq \mathbf{0} \text{ and } \mathbf{y} \geq \mathbf{0}. \quad (5.3.4)$$

The reader will note parenthetically that this definition of a saddle point agrees with the analogous concept in game theory (Section 1.4). The relation between solutions to Problem I and saddle points of  $\phi(\mathbf{x}, \mathbf{y})$  is clarified by Theorem 5.3.1 below.

The following lemma is of basic importance in the proofs of Theorems 5.3.1 and 5.4.1.

►LEMMA 5.3.1. Let  $S$  be a closed, convex, polyhedral set in Euclidean  $(m+n)$ -dimensional vector space ( $m \geq 1$ ) which satisfies the following two properties: (i) if

$$\mathbf{z} = \langle x_1, \dots, x_n, y_1, \dots, y_m \rangle = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

is in  $S$  and  $x_i \geq 0$  ( $i = 1, \dots, n$ ), then  $\mathbf{y} \geq \mathbf{0}$  [i.e., if  $\mathbf{y} \neq \mathbf{0}$ , then  $\mathbf{y}$  cannot have all components  $y_j$  ( $1 \leq j \leq m$ ) nonnegative]; (ii)  $S$  contains at least one point

$$\begin{pmatrix} \mathbf{x}^0 \\ \mathbf{y}^0 \end{pmatrix}$$

for which  $\mathbf{x}^0 \geq \mathbf{0}$ ,  $\mathbf{y}^0 \geq \mathbf{0}$  [by hypothesis (i)  $\mathbf{y}^0$  is necessarily identically zero]. Then there are vectors  $\mathbf{u}$  (of  $n$  components) and  $\mathbf{v}$  (of  $m$  components) such that

$$(1) \quad \mathbf{u}^0 \geq \mathbf{0}, \mathbf{v}^0 \gg \mathbf{0};$$

$$(2) \quad (\mathbf{u}^0, \mathbf{x}) + (\mathbf{v}^0, \mathbf{y}) \leq 0 \quad \text{for all } \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \text{ in } S.$$

*Remark.* The following statement of the geometric version of this lemma may help clarify its meaning. If  $S$  contains no point of the positive orthant except on the face  $\mathbf{y} = \mathbf{0}$  and does have a point on this face, then  $S$  can be separated by a hyperplane from the interior of the positive orthant and strictly from the face  $\mathbf{x} = \mathbf{0}$ . This lemma can be deduced as an easy con-

sequence of Theorem 3.1.1. However, in order to keep the content of this section independent of all previous advanced discussions, we present the following additional proof, which relies heavily on the duality theory of convex cones (see Section B.3 of the appendix).

*Proof.* The set  $S$  is first enlarged as follows. Define

$$T = \left\{ \begin{pmatrix} \xi \\ \eta \end{pmatrix} \mid \begin{pmatrix} \xi \\ \eta \end{pmatrix} \leq \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{for some } \begin{pmatrix} x \\ y \end{pmatrix} \in S \right\}.$$

(The inequality in the definition of  $T$  is required to hold for each coordinate.) It is easily verified that  $T$  is a closed, convex, polyhedral set which, by assumption (ii), necessarily contains the negative orthant.

Let  $T^*$  denote the dual polyhedral cone to  $T$  (see Section B.3). Explicitly, we define

$$T^* = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mid (\alpha, \xi) + (\beta, \eta) \leq 0 \quad \text{for } \begin{pmatrix} \xi \\ \eta \end{pmatrix} \in T \right\}.$$

The set  $T^*$  may then be characterized as

$$T^* = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mid \alpha \geq 0, \beta \geq 0, (\alpha, x) + (\beta, y) \leq 0 \quad \text{for all } \begin{pmatrix} x \\ y \end{pmatrix} \in S \right\}.$$

The conditions  $\alpha \geq 0, \beta \geq 0$  are necessary restrictions, since the cone  $T$  includes the negative orthant. The second reduction of the defining relations of  $T^*$  follows from the definition of  $T$  and the fact that  $\alpha$  and  $\beta$  are required to be nonnegative vectors.

To complete the proof of the lemma it is sufficient to show that for each fixed component  $j$  ( $1 \leq j \leq m$ ),  $T^*$  must intersect the set

$$H_j = \left\{ \begin{pmatrix} u \\ v \end{pmatrix} \mid u \geq 0, v \geq 0, v_j > 0 \right\}.$$

For in that case an appropriate convex combination of vectors of  $H_j$  may be constructed which fits the conclusion of the lemma. Now, to show that  $T^*$  intersects  $H_j$ , suppose to the contrary that  $T^* \subset K_{j_0}$ , where

$$K_{j_0} = \left\{ \begin{pmatrix} u \\ v \end{pmatrix} \mid u \geq 0, v \geq 0, v_{j_0} = 0 \right\}.$$

From the duality theory of convex cones, we deduce that

$$\begin{aligned} T^{**} \supset K_{j_0}^* = \left\{ \begin{pmatrix} \xi \\ \eta \end{pmatrix} \mid \xi \leq 0, \eta_1 \leq 0, \dots, \eta_{j_0-1} \leq 0, \right. \\ \left. \eta_{j_0+1} \leq 0, \dots, \eta_m \leq 0 \quad \text{and } \eta_{j_0} \text{ arbitrary} \right\}, \quad (5.3.5) \end{aligned}$$

where

$$T^{**} = \text{the cone with vertex at the origin spanned by } T, \quad (5.3.6)$$

since  $T$  is a closed, convex, polyhedral set which contains the negative orthant. Consider the vector  $\mathbf{c} = \begin{pmatrix} \mathbf{0} \\ \delta \end{pmatrix}$  whose first  $n$  coordinates are zero and whose last  $m$  coordinates are zero except for the  $j_0$ th coordinate, which has value  $\delta > 0$ . Obviously  $\mathbf{c}$  belongs to  $K_{j_0}^*$ . By virtue of the relationship of  $T$  and  $T^{**}$ , we infer the existence of  $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix}$  in  $S$  and  $\lambda \geq 0$  such that

$$\lambda x_i^* \geq 0 \quad (i = 1, \dots, n),$$

$$\lambda y_j^* \geq 0 \quad (j = 1, \dots, j_0 - 1, j_0 + 1, \dots, m),$$

$$\lambda y_{j_0}^* \geq \delta > 0.$$

The last inequality requires  $\lambda > 0$ ; hence  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mathbf{y}^* > \mathbf{0}$ , contrary to the hypothesis of the lemma. The proof of the lemma is complete.

►THEOREM 5.3.1. A vector  $\bar{\mathbf{x}}$  is a solution to the linear programming Problem I if and only if there exists a vector  $\bar{\mathbf{y}}$  such that  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of the Lagrangian form  $\phi(\mathbf{x}, \mathbf{y})$ .

*Proof.* We shall first prove that if  $\bar{\mathbf{x}}$  is a solution to Problem I, then there exists  $\bar{\mathbf{y}}$  such that  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{y})$ . Consider in the  $(m + 1)$ -dimensional vector space a set  $S$  defined by

$$S = \left\{ \begin{pmatrix} \mathbf{b} - \mathbf{Ax} \\ (\mathbf{c}, \mathbf{x} - \bar{\mathbf{x}}) \end{pmatrix} \middle| \mathbf{x} \geq \mathbf{0} \right\}.$$

A vector of  $S$  is an  $(m + 1)$ -tuple whose first  $m$  components constitute the image  $\mathbf{b} - \mathbf{Ax}$  of some nonnegative vector  $\mathbf{x}$ , and whose last component is the real number  $(\mathbf{c}, \mathbf{x} - \bar{\mathbf{x}})$  for the same  $\mathbf{x}$ . Obviously,  $S$  is a polyhedral convex set; and  $\mathbf{b} - \mathbf{Ax} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}$  implies  $(\mathbf{c}, \mathbf{x}) - (\mathbf{c}, \bar{\mathbf{x}}) \leq 0$ , since  $\bar{\mathbf{x}}$  is a solution of Problem I.

Moreover, substituting  $\bar{\mathbf{x}}$  for  $\mathbf{x}$  in the definition of  $S$  produces a nonnegative vector in  $S$ . Hence all the requirements of the hypothesis of Lemma 5.3.1 on  $S$  are met, and therefore there exists a vector  $\bar{\mathbf{y}}$  with  $m$  components such that

$$\bar{\mathbf{y}} \geq \mathbf{0} \quad (5.3.7)$$

and

$$(\mathbf{c}, \mathbf{x}) - (\mathbf{c}, \bar{\mathbf{x}}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{Ax}) \leq 0 \quad \text{for all } \mathbf{x} \geq \mathbf{0}. \quad (5.3.8)$$

Putting  $\mathbf{x} = \bar{\mathbf{x}}$  in (5.3.8) and recalling that

$$\mathbf{A}\bar{\mathbf{x}} \leq \mathbf{b}, \quad (5.3.9)$$

we have

$$(\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = 0 \quad \text{and} \quad (\mathbf{y}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) \geq 0 \quad \text{for } \mathbf{y} \geq \mathbf{0}. \quad (5.3.10)$$

Equations (5.3.7–10) together show that  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{y})$ .

Next we shall show that if  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{y})$ , then  $\bar{\mathbf{x}}$  is a solution to Problem I. The latter part of (5.3.4) can be written as

$$(\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) \leq (\mathbf{y}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) \quad \text{for all } \mathbf{y} \geq \mathbf{0}. \quad (5.3.11)$$

By choosing the appropriate coordinate of  $\mathbf{y}$  sufficiently large, we may deduce that

$$\mathbf{b} - \mathbf{A}\bar{\mathbf{x}} \geq \mathbf{0}, \quad \bar{\mathbf{x}} \geq \mathbf{0}. \quad (5.3.12)$$

Hence  $(\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) \geq 0$ . On the other hand, if we insert  $\mathbf{y} = \mathbf{0}$  in (5.3.11), it follows that

$$(\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) = 0. \quad (5.3.13)$$

The first part of (5.3.4) is written as

$$(\mathbf{c}, \mathbf{x}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\mathbf{x}) \leq (\mathbf{c}, \bar{\mathbf{x}}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) \quad \text{for all } \mathbf{x} \geq \mathbf{0}, \quad (5.3.14)$$

which by (5.3.13) is

$$(\mathbf{c}, \mathbf{x}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\mathbf{x}) \leq (\mathbf{c}, \bar{\mathbf{x}}) \quad \text{for all } \mathbf{x} \geq \mathbf{0}. \quad (5.3.15)$$

Therefore, for any  $\mathbf{x} \geq \mathbf{0}$  such that  $\mathbf{b} - \mathbf{A}\mathbf{x} \geq \mathbf{0}$ , we have

$$(\mathbf{c}, \mathbf{x}) \leq (\mathbf{c}, \bar{\mathbf{x}});$$

that is,  $\bar{\mathbf{x}}$  is a solution to Problem I.

**5.4 The principal theorems of linear programming (continued).** With the aid of Theorem 5.3.1 we are now prepared to establish, under very general conditions, the principal existence theorem for solutions to Problems I and II. Let us first consider the following duality theorem, which describes the interrelationships between solutions to Problems I and II.

► **THEOREM 5.4.1 (Duality Theorem).** If Problem I has a solution  $\bar{\mathbf{x}}$ , then Problem II has a solution  $\bar{\mathbf{y}}$  and

$$(\mathbf{c}, \bar{\mathbf{x}}) = (\bar{\mathbf{y}}, \mathbf{b}). \quad (5.4.1)$$

*Proof.* If  $\bar{\mathbf{x}}$  is a solution to Problem I, then by Theorem 5.3.1 there is a vector  $\bar{\mathbf{y}}$  such that  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of the Lagrangian form  $\phi(\mathbf{x}, \mathbf{y})$ , where

$$\begin{aligned}\phi(\mathbf{x}, \mathbf{y}) &= (\mathbf{c}, \mathbf{x}) + (\mathbf{y}, \mathbf{b} - \mathbf{Ax}) = -\{(\mathbf{y}, -\mathbf{b}) + (\mathbf{yA} - \mathbf{c}, \mathbf{x})\} \\ &= -\psi(\mathbf{y}, \mathbf{x}).\end{aligned}$$

Therefore

$$\psi(\mathbf{y}, \bar{\mathbf{x}}) \leq \psi(\bar{\mathbf{y}}, \bar{\mathbf{x}}) \leq \psi(\bar{\mathbf{y}}, \mathbf{x}), \quad (5.4.2)$$

and hence  $\{\bar{\mathbf{y}}, \bar{\mathbf{x}}\}$  is a saddle point of  $\psi(\mathbf{y}, \mathbf{x}) = (\mathbf{y}, -\mathbf{b}) + (\mathbf{yA} - \mathbf{c}, \mathbf{x})$ . Since  $\psi(\mathbf{y}, \mathbf{x})$  is the Lagrangian form of Problem II,  $\bar{\mathbf{y}}$  is a solution to Problem II, again by Theorem 5.3.1.

Since

$$(\bar{\mathbf{y}}, \mathbf{b} - \mathbf{Ax}) = 0 \quad \text{and} \quad (\bar{\mathbf{yA}} - \mathbf{c}, \bar{\mathbf{x}}) = 0 \quad (5.4.3)$$

(see 5.3.10), we have

$$(\bar{\mathbf{y}}, \mathbf{b}) = (\bar{\mathbf{y}}, \mathbf{Ax}) = (\bar{\mathbf{yA}}, \bar{\mathbf{x}}) = (\mathbf{c}, \bar{\mathbf{x}}).$$

► **THEOREM 5.4.2 (Existence Theorem).** If both Problem I and Problem II have feasible vectors, then both have solutions. All pairs of solutions  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{y}}$  satisfy  $(\mathbf{c}, \bar{\mathbf{x}}) = (\bar{\mathbf{y}}, \mathbf{b})$ .

*Proof.* Let  $\mathbf{y}^0$  be a feasible vector for Problem II; i.e.,  $\mathbf{y}^0 \geq \mathbf{0}$  and  $\mathbf{y}^0 \mathbf{A} \geq \mathbf{c}$ . Since the feasible set of Problem I is nonvoid, take any feasible vector  $\mathbf{x}$ . Then

$$(\mathbf{c}, \mathbf{x}) \leq (\mathbf{y}^0 \mathbf{A}, \mathbf{x}) \leq (\mathbf{y}^0, \mathbf{b}). \quad (5.4.4)$$

Therefore, the set  $F = \{(\mathbf{c}, \mathbf{x}) | \mathbf{x} \text{ feasible for Problem I}\}$  is bounded above, and since the feasible set  $F$  is a closed, polyhedral, convex set, the maximum of  $(\mathbf{c}, \mathbf{x})$  is achieved.

The remaining assertions of the theorem follow by applying Theorem 5.4.1.

► **COROLLARY 5.4.1.** If  $\mathbf{x}^*$  and  $\mathbf{y}^*$  are feasible for Problems I and II, respectively, and  $(\mathbf{c}, \mathbf{x}^*) = (\mathbf{y}^*, \mathbf{b})$ , then  $\mathbf{x}^*, \mathbf{y}^*$  constitute a pair of solutions, and conversely.

*Proof.* It follows from (5.4.4) that if  $\mathbf{x}$  and  $\mathbf{y}$  are any pair of feasible vectors for Problems I and II, respectively, then

$$(\mathbf{c}, \mathbf{x}) \leq (\mathbf{y}, \mathbf{b}). \quad (5.4.5)$$

For the special choice of  $\mathbf{x}$  and  $\mathbf{y}^*$ , we find

$$(\mathbf{c}, \mathbf{x}) \leq (\mathbf{y}^*, \mathbf{b}) = (\mathbf{c}, \mathbf{x}^*),$$

and therefore  $\mathbf{x}^*$  is a solution to Problem I.

Analogously,  $\mathbf{y}^*$  is a solution to Problem II. The converse proposition is the substance of Theorem 5.4.1.

The corollary provides a simple computational criterion for checking whether pairs of feasible vectors are solutions. Moreover, by virtue of (5.4.5) feasible vectors can be used to generate estimates and bounds for Problems I and II. These applications are explored in Chapter 6, where we develop computing methods for solving linear programming problems.

Because of its economic significance, a property of solutions  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  which is implicitly expressed in (5.4.3) is now stated as a theorem.

► **THEOREM 5.4.3.** If  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{y}}$  are solutions to Problems I and II, respectively, then

- (a)  $(\mathbf{A}\bar{\mathbf{x}})_j < b_j$  implies  $\bar{y}_j = 0$ ,
- (b)  $(\bar{\mathbf{y}}\mathbf{A})_i > c_i$  implies  $\bar{x}_i = 0$ .

These characteristics of solutions are clearly valuable in any computational algorithm, but their basic importance derives from their economic interpretation.

Consider, for example, the activity analysis model and its dual (as formulated in Sections 5.1 and 5.2). In this context part (a) of the theorem states that if  $\bar{\mathbf{x}}$  represents an optimal program of effort which does not utilize all of the  $j$ th resource, then such a resource is overabundant and has no money value in a competitive market at equilibrium. Part (b) asserts that if the value of the  $i$ th activity is less than its operating cost, then that activity is not undertaken (its level is zero). Both conditions make sense economically. The second says in effect that no activity which loses money will be undertaken. The first may be considered a corollary of the law of supply and demand.

**\*5.5 Connections between linear programming problems and game theory.** In the last two sections we established the existence of solutions to Problems I and II, and described how such solutions are related. In doing so, we introduced a saddle-point or game-type problem (the Lagrangian).

We now establish a connection between solutions of Problems I and II and the solution of the matrix game

$$\mathbf{B} = \begin{array}{c|ccc} & n & m & 1 \\ \hline n & 0 & -\mathbf{A}' & \mathbf{c} \\ m & \mathbf{A} & 0 & -\mathbf{b} \\ \hline 1 & -\mathbf{c} & \mathbf{b} & 0 \end{array}.$$

---

\* Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

The numbers  $m$ ,  $n$ , and 1 outside the matrix denote the number of rows and columns in the various submatrices of  $B$ .  $B$  is a skew-symmetric matrix; hence the value of the game is zero. Moreover, the sets of strategies for both players are identical (see Section 2.5). A strategy for either player will be denoted by  $\langle \mathbf{x}, \mathbf{y}, \lambda \rangle = \langle x_1, \dots, x_n, y_1, \dots, y_m, \lambda \rangle$ .

► **THEOREM 5.5.1.** If the symmetric game  $\mathbf{B}$  has a solution for which  $\lambda > 0$ , then both linear programming problems have solutions, and conversely.

*Proof.* Let  $\mathbf{z} = \langle x_1, \dots, x_n, y_1, \dots, y_m, \lambda \rangle$  be an optimal strategy for the matrix game  $\mathbf{B}$  with  $\lambda > 0$ . Since the value  $v$  of the game is zero, we obtain  $\mathbf{Bz} \leq 0$ , or

$$-\mathbf{yA} + \mathbf{c}\lambda \leq \mathbf{0}, \quad \mathbf{Ax} - \mathbf{b}\lambda \leq \mathbf{0}, \quad -(\mathbf{c}, \mathbf{x}) + (\mathbf{b}, \mathbf{y}) \leq 0.$$

Since  $\lambda > 0$ , it follows from Lemma 2.1.2 that equality holds in the last relation. Therefore

$$\frac{\mathbf{y}}{\lambda} \mathbf{A} \geq \mathbf{c}, \quad \mathbf{A} \frac{\mathbf{x}}{\lambda} \leq \mathbf{b}, \quad \left( \mathbf{c}, \frac{\mathbf{x}}{\lambda} \right) = \left( \mathbf{b}, \frac{\mathbf{y}}{\lambda} \right). \quad (5.5.1)$$

Let  $\mathbf{y}' = \mathbf{y}/\lambda$ , and  $\mathbf{x}' = \mathbf{x}/\lambda$ . These vectors are feasible by (5.5.1). Moreover, for any feasible  $\mathbf{x}$  in  $X$  (the set of all feasible vectors),

$$(\mathbf{c}, \mathbf{x}) \leq (\mathbf{y}'\mathbf{A}, \mathbf{x}) = (\mathbf{y}', \mathbf{Ax}) \leq (\mathbf{y}', \mathbf{b}) = (\mathbf{c}, \mathbf{x}').$$

Hence

$$\max_{\mathbf{x} \in X} (\mathbf{c}, \mathbf{x}) = (\mathbf{c}, \mathbf{x}').$$

Similarly,

$$\min_{\mathbf{y} \in Y} (\mathbf{y}, \mathbf{b}) = (\mathbf{y}', \mathbf{b}).$$

Consequently,  $\mathbf{x}'$  and  $\mathbf{y}'$  are solutions to Problems I and II, respectively.

Conversely, let  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  and  $\mathbf{y} = \langle y_1, \dots, y_m \rangle$  be two solutions to the linear programming Problems I and II, respectively. Set

$$\lambda = \frac{1}{1 + \sum x_i + \sum y_i},$$

and let  $\bar{x}_i = \lambda x_i$ ,  $\bar{y}_i = \lambda y_i$ ,  $\mathbf{z} = \langle \bar{x}_1, \dots, \bar{x}_n, \bar{y}_1, \dots, \bar{y}_m, \lambda \rangle = \langle \bar{\mathbf{x}}, \bar{\mathbf{y}}, \lambda \rangle$ . Clearly,  $\sum \bar{z}_i = 1$ . We now verify that  $\bar{\mathbf{z}}$  is a solution to the symmetric game  $\mathbf{B}$ :

$$\mathbf{yA} \geq \mathbf{c} \rightarrow \lambda \mathbf{yA} \geq \lambda \mathbf{c} \rightarrow -\bar{\mathbf{yA}} + \lambda \mathbf{c} \leq \mathbf{0},$$

$$\mathbf{Ax} \leq \mathbf{b} \rightarrow \mathbf{A}\lambda \mathbf{x} \leq \lambda \mathbf{b} \rightarrow \mathbf{A}\bar{\mathbf{x}} - \lambda \mathbf{b} \leq \mathbf{0},$$

and

$$(\mathbf{c}, \mathbf{x}) = (\mathbf{y}, \mathbf{b})$$

by Theorem 5.4.1. Combining, we get  $\mathbf{B}\bar{\mathbf{z}} \leq \mathbf{0}$ ; hence  $\bar{\mathbf{z}}$  is a solution to the game.

By virtue of Theorem 5.5.1, useful ideas from both game theory and linear programming can be applied to problems in either discipline.

**\*5.6 Extensions of the duality theorem.** Extensions of Problems I and II may be formulated as follows: Let  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  denote specified matrices of order  $r \times n$ ,  $r \times m$ , and  $s \times n$ , respectively, and let  $\mathbf{D}$  describe an arbitrary matrix of order  $s \times m$ . The generic vectors  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{u}$ , and  $\mathbf{v}$  lie respectively in  $E^n$ ,  $E^m$ ,  $E^r$ , and  $E^s$ .

*Problem A.* From among all matrices  $\mathbf{D}$  which satisfy  $\mathbf{Dy} \leq \mathbf{Cx}$ , where  $\mathbf{Ax} \leq \mathbf{By}$  for some  $\mathbf{x} \geq \mathbf{0}$  and  $\mathbf{y} \gg \mathbf{0}$ , find a maximal matrix  $\mathbf{D}$ .

A matrix  $\mathbf{D}$  is said to be maximal if there exists no  $\mathbf{D}'$  which fulfills the conditions given above for some  $\mathbf{x} \geq \mathbf{0}$  and  $\mathbf{y} \gg \mathbf{0}$  and  $\mathbf{D}' > \mathbf{D}$ ; that is, each component of  $\mathbf{D}'$  is at least as large as the corresponding component of  $\mathbf{D}$  and at least one component is larger.

*Problem B.* Find a minimal matrix  $\mathbf{D}$  satisfying  $\mathbf{vD} \geq \mathbf{uB}$  where  $\mathbf{uA} \geq \mathbf{vC}$  for some  $\mathbf{v} \gg \mathbf{0}$  and  $\mathbf{u} \geq \mathbf{0}$ .

*Remark.* If  $\mathbf{A}$  is a general  $m \times n$  matrix,  $\mathbf{B} = (m \times 1)$ ,  $\mathbf{C} = (1 \times n)$ , and  $\mathbf{D} = (1 \times 1)$ , that is,  $\mathbf{D}$  is a real number, then Problem A is equivalent to Problem I.

There follow extensions of Theorems 5.4.1 and 5.4.2, respectively.

► **THEOREM 5.6.1.** If Problem A has a solution  $\mathbf{D}$ , then Problem B has  $\mathbf{D}$  as a solution. Conversely, if Problem B has a solution  $\Delta$ , then Problem A has  $\Delta$  as a solution.

► **THEOREM 5.6.2.** If there exist vectors  $\mathbf{x}_0 \geq \mathbf{0}$ ,  $\mathbf{y}_0 \gg \mathbf{0}$ ,  $\mathbf{u}_0 \geq \mathbf{0}$ , and  $\mathbf{v}_0 \gg \mathbf{0}$  such that  $\mathbf{Ax}_0 \leq \mathbf{By}_0$  and  $\mathbf{u}_0\mathbf{A} \geq \mathbf{v}_0\mathbf{C}$ , then Problem A and Problem B have identical sets of solutions.

*Proof of Theorem 5.6.1.* Let  $\mathbf{D}$  be a solution to Problem A. Consider in  $(r + s)$ -dimensional space a set  $S$  defined by

$$S = \left\{ \begin{pmatrix} \mathbf{By} - \mathbf{Ax} \\ \mathbf{Cx} - \mathbf{Dy} \end{pmatrix} \middle| \mathbf{x} \geq \mathbf{0}, \mathbf{y} \geq \epsilon \right\},$$

where  $\epsilon$  is the vector all of whose components are equal to  $\epsilon$  (with  $\epsilon > 0$

and fixed). The set  $S$  is a closed, convex, polyhedral set such that

$$\mathbf{By} - \mathbf{Ax} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \mathbf{y} \geq \epsilon \quad \text{implies} \quad \mathbf{Cx} - \mathbf{Dy} \gg \mathbf{0}. \quad (5.6.1)$$

Otherwise,  $\mathbf{Cx} \geq \mathbf{Dy}$  with strict inequality for some component, and one of the elements of  $\mathbf{D}$  can be increased without affecting the inequality  $\mathbf{Cx} \geq \mathbf{Dy}$ ; this contradicts the hypothesis that  $\mathbf{D}$  is a maximal matrix of Problem A.

Since  $\mathbf{D}$  is maximal to Problem A, there are vectors  $\mathbf{x}^0 \geq \mathbf{0}$  and  $\mathbf{y}^0 \gg \mathbf{0}$  such that

$$\mathbf{Dy}^0 = \mathbf{Cx}^0 \quad (5.6.2)$$

and

$$\mathbf{Ax}^0 \leq \mathbf{By}^0. \quad (5.6.3)$$

We select  $\epsilon > 0$  sufficiently small so that  $\mathbf{y}^0 \geq \epsilon$  and hence  $\mathbf{x}^0, \mathbf{y}^0$  qualify in the construction of  $S$ . For this  $\epsilon$ , we see that  $S$  contains a vector none of whose coordinates are negative. Thus, conditions (i) and (ii) of Lemma 5.3.1 are satisfied for the set  $S$ .

By Lemma 5.3.1, there exist vectors  $\mathbf{u}^0$  and  $\mathbf{v}^0$  such that  $\mathbf{u}^0 \geq \mathbf{0}, \mathbf{v}^0 \gg \mathbf{0}$ , and

$$(\mathbf{u}^0, \mathbf{By} - \mathbf{Ax}) + (\mathbf{v}^0, \mathbf{Cx} - \mathbf{Dy}) \leq 0 \quad \text{for all } \mathbf{x} \geq \mathbf{0}, \mathbf{y} \gg \epsilon. \quad (5.6.4)$$

Rewriting (5.6.4) gives us

$$(\mathbf{u}^0 \mathbf{B} - \mathbf{v}^0 \mathbf{D}, \mathbf{y}) + (\mathbf{v}^0 \mathbf{C} - \mathbf{u}^0 \mathbf{A}, \mathbf{x}) \leq 0 \quad \text{for all } \mathbf{x} \geq \mathbf{0}, \mathbf{y} \gg \epsilon. \quad (5.6.5)$$

For  $\mathbf{x} = \mathbf{0}$  in (5.6.5) we get  $(\mathbf{u}^0 \mathbf{B} - \mathbf{v}^0 \mathbf{D}, \mathbf{y}) \leq 0$  for all  $\mathbf{y} \gg \epsilon$ . Hence

$$\mathbf{u}^0 \mathbf{B} \leq \mathbf{v}^0 \mathbf{D}. \quad (5.6.6)$$

Since (5.6.5) holds for any  $\mathbf{x} \geq \mathbf{0}, \mathbf{y} \gg \epsilon$ , we have

$$\mathbf{v}^0 \mathbf{C} \leq \mathbf{u}^0 \mathbf{A}. \quad (5.6.7)$$

Equations (5.6.6) and (5.6.7) show that  $\mathbf{D}$  is a feasible matrix for Problem B.

We shall now prove that  $\mathbf{D}$  is minimal for Problem B. Assume that  $\mathbf{D}$  is not minimal; then there would be a matrix  $\mathbf{A}$  such that

$$\mathbf{A} < \mathbf{D}, \quad (5.6.8)$$

$$\mathbf{vA} \geq \mathbf{uB}, \quad (5.6.9)$$

$$\mathbf{uA} \geq \mathbf{vC}, \quad (5.6.10)$$

for some  $\mathbf{u} \geq \mathbf{0}$  and  $\mathbf{v} \gg \mathbf{0}$ .

It follows that

$$\begin{aligned}
 (\mathbf{v}, \mathbf{Dy}^0) &> (\mathbf{v}, \Delta \mathbf{y}^0) \geq (\mathbf{u}, \mathbf{By}^0) && \text{by (5.6.8) and (5.6.9)} \\
 &\geq (\mathbf{u}, \mathbf{Ax}^0) && \text{by (5.6.3)} \\
 &\geq (\mathbf{v}, \mathbf{Cx}^0) && \text{by (5.6.10)} \\
 &\geq (\mathbf{v}, \mathbf{Dy}^0) && \text{by (5.6.2),}
 \end{aligned}$$

which is a contradiction. Therefore  $\mathbf{D}$  is minimal for Problem B. The second half of the theorem is proved analogously.

*Proof of Theorem 5.6.2.* Let  $\mathbf{By}_0 = \mathbf{b}$  and  $\mathbf{v}_0 \mathbf{C} = \mathbf{c}$ , and consider the symmetric game with matrix

$$\begin{vmatrix} \mathbf{0} & -\mathbf{A}' & +\mathbf{c} \\ \mathbf{A} & \mathbf{0} & -\mathbf{b} \\ -\mathbf{c} & +\mathbf{b} & \mathbf{0} \end{vmatrix}.$$

Let  $\langle \mathbf{x}, \mathbf{u}, \lambda \rangle$  denote an optimal strategy for Player I. Then  $(\mathbf{c}, \mathbf{x}) - (\mathbf{b}, \mathbf{u}) \geq 0$ . If  $\lambda = 0$ , then  $\mathbf{uA} \geq \mathbf{0}$  and  $\mathbf{Ax} \leq \mathbf{0}$ , and consequently

$$0 \leq (\mathbf{uA}, \mathbf{x}_0) = (\mathbf{u}, \mathbf{Ax}_0) \leq (\mathbf{u}, \mathbf{By}_0) = (\mathbf{u}, \mathbf{b}).$$

Similarly,  $\mathbf{Ax} \leq \mathbf{0}$  leads to the conclusion that  $(\mathbf{c}, \mathbf{x}) \leq 0$ . On the other hand, if  $\lambda > 0$ , then by Lemma 2.1.2  $(\mathbf{c}, \mathbf{x}) - (\mathbf{b}, \mathbf{u}) = 0$ . Thus we see that under all circumstances we have  $(\mathbf{c}, \mathbf{x}) - (\mathbf{b}, \mathbf{u}) = 0$ .

By Theorem 3.1.1, there must exist an optimal strategy with  $\lambda > 0$ . In any such strategy,  $(\mathbf{c}, \mathbf{x}) = (\mathbf{u}, \mathbf{b})$ ,  $\mathbf{uA} \geq \lambda \mathbf{c}$ , and  $\lambda \mathbf{b} \geq \mathbf{Ax}$ . The vectors  $\mathbf{u}$  and  $\mathbf{x}$  may presumably be zero. Let  $\bar{\mathbf{u}} = \mathbf{u}/\lambda$  and  $\bar{\mathbf{x}} = \mathbf{x}/\lambda$ . Then  $\bar{\mathbf{u}}\mathbf{A} \geq \mathbf{c} = \mathbf{v}_0 \mathbf{C}$ ;  $\bar{\mathbf{A}}\bar{\mathbf{x}} \leq \mathbf{b} = \mathbf{By}_0$ ; and  $(\bar{\mathbf{u}}, \mathbf{b}) = (\mathbf{c}, \bar{\mathbf{x}})$ , or  $(\bar{\mathbf{u}}, \mathbf{By}_0) = (\mathbf{v}_0 \mathbf{C}, \bar{\mathbf{x}})$ , or equivalently  $(\bar{\mathbf{u}}\mathbf{B}, \mathbf{y}_0) = (\mathbf{v}_0, \mathbf{C}\bar{\mathbf{x}})$ . If  $(\bar{\mathbf{u}}\mathbf{B}, \mathbf{y}_0) \neq 0$ , define

$$\mathbf{Dy} = \frac{(\bar{\mathbf{u}}\mathbf{B}, \mathbf{y})\mathbf{C}\bar{\mathbf{x}}}{(\bar{\mathbf{u}}\mathbf{B}, \mathbf{y}_0)};$$

otherwise define

$$\mathbf{Dy} = \frac{(\bar{\mathbf{u}}\mathbf{B}, \mathbf{y})\mathbf{1}}{(\mathbf{v}_0, \mathbf{1})} + \frac{(\mathbf{1}, \mathbf{y})\mathbf{C}\bar{\mathbf{x}}}{(\mathbf{1}, \mathbf{y}_0)},$$

where  $\mathbf{1}$  is the vector  $\mathbf{1} = \langle 1, 1, \dots, 1 \rangle$ . It is readily verified that  $\mathbf{Dy}_0 = \mathbf{C}\bar{\mathbf{x}}$  and  $\mathbf{v}_0 \mathbf{D} = \bar{\mathbf{u}}\mathbf{B}$ , and also that the inequalities  $\bar{\mathbf{A}}\bar{\mathbf{x}} \leq \mathbf{By}_0$  and  $\bar{\mathbf{u}}\mathbf{A} \geq \mathbf{v}_0 \mathbf{C}$  are valid. Thus the matrix  $\mathbf{D}$  constructed above satisfies the requirements of Problems A and B. It remains only to show that  $\mathbf{D}$  is the solution to the problems. Suppose  $\Delta > \mathbf{D}$  has the property that  $\mathbf{Ax}' \leq \mathbf{By}'$  and  $\mathbf{Cx}' \geq \Delta \mathbf{y}'$  where  $\mathbf{y}' \gg \mathbf{0}$  and  $\mathbf{x}' \geq \mathbf{0}$ . This last fact implies that  $\mathbf{Cx}' \geq \Delta \mathbf{y}' >$

$Dy'$ , with inequality for at least one component. The properties of  $\{\bar{u}, v_0\}$  imply that

$$\begin{vmatrix} \xi \\ \eta \end{vmatrix} = \begin{vmatrix} -A' & C' \\ B' & -D' \end{vmatrix} \begin{vmatrix} \bar{u} \\ v_0 \end{vmatrix} \leq 0. \quad (5.6.11)$$

Consequently, the inner product of the nonnegative vector  $\{x', y'\}$  and  $\{\xi, \eta\}$  gives a nonpositive result. However,

$$\begin{vmatrix} \alpha \\ \beta \end{vmatrix} = \begin{vmatrix} -A & B \\ C & -D \end{vmatrix} \begin{vmatrix} x' \\ y' \end{vmatrix} \geq 0, \quad (5.6.12)$$

with inequality somewhere in the last set of components corresponding to  $\beta$ . Thus, since  $v_0 \gg 0$ , the inner product of  $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$  and  $\{\bar{u}, v_0\}$  is strictly positive. But both inner products yield the same formula, one obtained from the other by rearrangement. Hence we obtain a contradiction. The proof of the theorem is complete.

**5.7 Warehouse problem.** Sections 5.7 through 5.12 are devoted to analytical studies of specific linear programming models.

In this section we examine the warehouse storage problem introduced in Section 5.1. A warehouse of capacity  $C$  is used to stock merchandise. The initial amount of merchandise on hand is assumed to be  $A$ . It is desired to determine how much stock to buy and sell during  $n$  periods of time so as to maximize profit. Buying takes place at the end of each period. Let  $x_i$  be the amount of goods bought at the end of the  $i$ th period,  $c_i$  the cost of the goods at the end of the  $i$ th period,  $y_i$  the amount of stock sold during the  $i$ th period, and  $b_i$  the selling price during the  $i$ th period. We wish to maximize

$$\pi = \sum_{i=1}^n [b_i y_i - c_i x_i].$$

There are two constraints for the problem: (1) no more may be bought than the warehouse can hold, and (2) no more can be sold than is available in stock. Expressed mathematically, these constraints are

$$A + \sum_{j=1}^i (x_j - y_j) \leq C \quad (i = 1, 2, \dots, n) \quad (5.7.1)$$

and

$$y_{i+1} \leq A + \sum_{j=1}^i (x_j - y_j) \quad (i = 0, 1, \dots, n-1). \quad (5.7.2)$$

In addition, there are the usual constraints: that is,  $x_i \geq 0$  and  $y_i \geq 0$  for  $i = 1, 2, \dots, n$ .

We first solve the dual problem, after which a solution of the original problem is immediate. The dual problem is to minimize

$$(C - A) \sum_{i=1}^n t_i + A \sum_{i=1}^n u_i,$$

where  $t_i \geq 0$ ,  $u_i \geq 0$ , and the corresponding constraints are

$$T_k - U_{k+1} \geq -c_k \quad (k = 1, 2, \dots, n), \quad -T_k + U_k \geq b_k \quad (k = 1, 2, \dots, n);$$

by definition,

$$T_k = \sum_{i=k}^n t_i, \quad U_k = \sum_{i=k}^n u_i, \quad \text{and} \quad U_{n+1} = 0.$$

The nonnegativity of  $t_i$  and  $u_i$  are equivalent to the conditions

$$T_{k-1} \geq T_k, \quad T_n \geq 0 \quad \text{and} \quad U_{k-1} \geq U_k, \quad U_n \geq 0. \quad (5.7.3)$$

We seek to minimize  $(C - A)T_1 + AU_1$  subject to these constraints.

It is clear from the constraining relations (5.7.3) that the minimum  $T_k$ 's and  $U_k$ 's are

$$\begin{aligned} T_n^* &= \max(-c_n, 0), & U_n^* &= \max(b_n + T_n^*, 0), \\ T_{n-1}^* &= \max(U_n^* - c_{n-1}, T_n^*), & & \\ U_{n-1}^* &= \max(b_{n-1} + T_{n-1}^*, U_n), & & \\ T_{n-2}^* &= \max(U_{n-1}^* - c_{n-2}, T_{n-1}^*), & \text{etc.}, & \end{aligned} \quad (5.7.4)$$

and finally the formula for  $T_1^*$  and  $U_1^*$  obtains. In short, if at any stage a value  $U_k(T_k)$  is chosen larger than the value  $U_k^*(T_k^*)$ , then the constraints imply that  $T_1$  and  $U_1$  exceed  $T_1^*$  and  $U_1^*$ , respectively. Consequently,  $T_1^*$  and  $U_1^*$  are the solutions to the dual problem.

It is clear from the formation rules that  $T_1^*$  and  $U_1^*$  are piecewise linear functions of  $b_i$  and  $c_i$ . Hence the terms of  $(C - A)T_1^* + AU_1^*$  may be rearranged into the form

$$(C - A)T_1^* + AU_1^* = \sum_i [b_i f_i(A, C) - c_i g_i(A, C)], \quad (5.7.5)$$

where  $f_i(A, C)$  and  $g_i(A, C)$  are functions of  $A$  and  $C$  only. The fact that  $f_i$  and  $g_i$  are nonnegative is clear by construction, since in (5.7.4) the coefficients of  $b_j$  are always nonnegative and the coefficients of  $c_j$  are nonpositive.

We show below that  $y_i^0 = f_i(A, C)$  and  $x_i^0 = g_i(A, C)$  are feasible vectors. It follows from Corollary 5.4.1 that the pair  $\mathbf{x}^0, \mathbf{y}^0$  as defined maximizes profit.

It remains only to verify that  $x_i^0$  and  $y_i^0$  ( $i = 1, 2, \dots, n$ ) are feasible. The recurrence relations (5.7.4) imply that one of the following two cases occurs:

*Case A.*

$$\begin{aligned} U_1^* &= b_1 - c_1 + b_2 - c_2 + \dots + b_i + T_{i+1}^*, \\ T_1^* &= -c_1 + b_2 - c_2 + \dots + b_i + T_{i+1}^*, \end{aligned} \quad (5.7.6)$$

for some  $1 < i \leq n$ , or

$$\begin{aligned} U_1^* &= b_1 + T_2^*, \\ T_1^* &= T_2^*. \end{aligned} \quad (5.7.6a)$$

*Case B.*

$$\begin{aligned} U_1^* &= b_1 - c_1 + b_2 - c_2 + \dots - c_j + U_{j+2}^*, \\ T_1^* &= -c_1 + b_2 - c_2 + \dots - c_j + U_{j+2}^*, \end{aligned} \quad (5.7.7)$$

for some  $1 < j \leq n - 1$  (if  $j = n - 1$ , then assign  $U_{j+2}^* = 0$ ); or

$$\begin{aligned} U_1^* &= U_2^*, \\ T_1^* &= -c_1 + U_2^*. \end{aligned} \quad (5.7.7a)$$

Another application yields

$$T_{i+1}^* = -c_{i+k_1} + b_{i+k_2} - c_{i+k_3} + \dots,$$

where  $1 < k_1 < k_2 < \dots$  (if  $i + k_1 > n$ , then  $T_{i+1}^* \equiv 0$ ). Similarly,

$$U_{j+2}^* = +b_{j+2+l_1} - c_{j+2+l_2} + \dots,$$

where  $0 < l_1 < l_2 < \dots$  (if  $j + 2 + l_1 > n - 1$ , then  $U_{j+2}^* = 0$ ). Once the form of  $U_1^*$  and  $T_1^*$  is known according to (5.7.6) or (5.7.7), it is easy to verify directly that the constraints (5.7.1) and (5.7.2) are satisfied.

**5.8 Optimal assignment problem.** There are  $m$  jobs to be done and  $m$  workers to do them; one worker is assigned to each job. The value of having worker  $j$  do job  $i$  is  $a_{ij}$  ( $a_{ij} > 0$ ). The problem is to determine the  $m$  assignments that will maximize the total value. An assignment of workers to jobs is denoted by the permutation symbol

$$S = \begin{pmatrix} 1, 2, \dots, m \\ j_1, j_2, \dots, j_m \end{pmatrix},$$

where worker  $j_1$  does job 1, worker  $j_2$  does job 2, etc. Hence the problem is to maximize

$$\sum_{i=1}^m a_{ij_i}$$

over the  $m!$  possible  $S$  obtained by making all possible permutations of  $1, 2, \dots, m$ , each permutation being an  $n$ -tuple  $\langle j_1, j_2, \dots, j_m \rangle$ .

An alternative way of describing the procedures is in terms of permutation matrices. A permutation matrix is an  $m \times m$  matrix each row and column of which contains one element of value 1 and zeros elsewhere. If  $\mathcal{P}$  is the set of all  $m \times m$  permutation matrices  $\mathbf{P} = \{p_{ij}\}$ , then the problem is to find  $\mathbf{P}^0$  such that

$$\max_{\mathcal{P}} \sum_{i,j} a_{ij} p_{ij} = \sum_{i,j} a_{ij} p_{ij}^0.$$

Von Neumann treats the assignment problem by reducing it to a  $2m \times m^2$  game. In most cases it will be easier to solve the game than to search for the maximum in the set of all permutation matrices. Before demonstrating the equivalence of the assignment problem to a game problem, we consider it as a problem in linear programming.

(1) Consider the set  $\mathfrak{X}$  of all  $m \times m$  doubly stochastic matrices. That is,  $\|x_{ij}\| \in \mathfrak{X}$  implies  $\sum_i x_{ij} = 1$  for all  $j$ ,  $\sum_j x_{ij} = 1$  for all  $i$ , and  $x_{ij} \geq 0$  for all  $i, j$ . Clearly,  $\mathfrak{X}$  is a convex polyhedral set, since it is the intersection of  $2m$  hyperplanes with the positive orthant of  $m^2$ -space.

►LEMMA 5.8.1. The matrices in the class  $\mathfrak{X}$  form a convex set, and the permutation matrices are its extreme points.

*Proof.* The convexity of  $\mathfrak{X}$  was noted above. Let  $\mathbf{X} = \{x_{ij}\}$  be an extreme point.

(i) At most  $2m - 1$  elements of  $\mathbf{X}$  are strictly positive.

In order to establish this assertion, we consider the following system of equations:

$$\begin{aligned} \sum_{j=1}^m z_{ij} &= 0 & (i = 1, 2, \dots, m), \\ \sum_{i=1}^m z_{ij} &= 0 & (j = 1, 2, \dots, m). \end{aligned} \tag{5.8.1}$$

The system contains  $2m$  equations in  $m^2$  variables, but there are at most  $2m - 1$  independent equations, since  $\sum_i (\sum_j z_{ij}) - \sum_j (\sum_i z_{ij}) = 0$ . The matrix  $\mathbf{A}$  of the system (5.8.1) is of the form

$$\mathbf{A} = \left( \begin{array}{cccc|c} & \overbrace{1 1 \cdots 1}^m & & & & \\ & 1 1 \cdots 1 & & & & \\ & & 1 1 \cdots 1 & & & \\ & & & \ddots & & \\ & & & & \ddots 1 1 \cdots 1 & \\ \hline 1 & 1 & 1 & \cdots 1 & & \\ 1 & 1 & 1 & & 1 & \\ \hline & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & & & \\ & & & 1 & 1 & \cdots 1 & 1 \end{array} \right) \quad (5.8.2)$$

(all entries are 0 except those indicated to be 1). We know that  $\mathbf{A}$  is of size  $2m \times m^2$  and has rank at most  $2m - 1$ .

Suppose that  $\mathbf{X} = \|x_{ij}\|$  has at least  $2m$  nonzero components. Whenever  $x_{ij} = 0$ , we assign to  $z_{ij}$  the value zero in the equations (5.8.1) and view (5.8.1) as a system of equations for the remaining variables. Since there are at least  $2m$  undetermined variables and the rank of  $\mathbf{A}$  is  $2m - 1$  (see Problem 6 of Chapter 6), there exists a nontrivial solution denoted by  $z_{ij}$ .

Choose  $\epsilon$  sufficiently small so that  $x_{ij} + \epsilon z_{ij} \geq 0$  and  $x_{ij} - \epsilon z_{ij} \geq 0$  for all  $i, j$ . Set  $\mathbf{X}_1 = \|x_{ij} + \epsilon z_{ij}\|$  and  $\mathbf{X}_2 = \|x_{ij} - \epsilon z_{ij}\|$ .

It is readily verified that  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are members of  $\pi$ .

But  $\mathbf{X} = \frac{1}{2}(\mathbf{X}_1 + \mathbf{X}_2)$ ; hence  $\mathbf{X}$  is not an extreme point.

(ii) If  $\mathbf{X}$  is an extreme point, then  $\mathbf{X}$  is a permutation matrix and conversely.

The proof of this statement is as follows. By (i), if  $\mathbf{X}$  is extreme, it can have at most  $2m - 1$  nonzero elements. Then there must be at least one row with only one nonzero element in it. Since the sum over this row is 1, the nonzero element must be 1. Moreover, this also makes this element the only nonzero element in its column.

But if  $\mathbf{X}$  is an extreme point of the class  $\mathfrak{X}$ , the matrix obtained by deleting the row and the column with one element is also an extreme point of the class of  $(m - 1) \times (m - 1)$  matrices of type  $\mathfrak{X}$ . Repeating the argument, we see that  $\mathbf{X}$  must have only one nonzero element in each row or column, and  $\mathbf{X}$  therefore represents a permutation matrix.

Conversely, any matrix that represents a permutation is easily seen to be an extreme point of  $\mathfrak{X}$ . The proof of the lemma is complete.

Since  $\Phi \subset \mathfrak{X}$ ,

$$\max_{x \in \mathfrak{X}} \sum a_{ij} x_{ij} \geq \max_{P \in \Phi} \sum a_{ij} p_{ij}.$$

But  $\|x_{ij}\| \in \mathfrak{X}$  is a finite convex combination of extreme points of  $\mathfrak{X}$ , so that (by Lemma 5.8.1)

$$\max_{x \in \mathfrak{X}} \sum a_{ij} x_{ij} = \max_{P \in \Phi} \sum a_{ij} p_{ij}.$$

Thus  $\mathbf{X}^0 \in \mathfrak{X}$  which maximizes  $\sum a_{ij} x_{ij}$  must be either an element of  $\Phi$  or a convex combination of elements of  $\Phi$  all of which maximize  $\sum a_{ij} p_{ij}$ . Therefore if we know  $\mathbf{X}^0$ , we can find a  $\mathbf{P}^0 \in \Phi$  which maximizes  $\sum a_{ij} p_{ij}$ . Consequently, it is sufficient in searching for a maximum to consider  $\mathfrak{X}$  instead of  $\Phi$ .

We further enlarge our domain of consideration to the set  $\mathcal{Y}$  of matrices  $\|y_{ij}\|$  which satisfy  $\sum_i y_{ij} \leq 1$  for all  $j$ ,  $\sum_j y_{ij} \leq 1$  for all  $i$ , and  $y_{ij} \geq 0$  for all  $i, j$ . It follows that

$$\max_{\mathcal{Y}} \sum a_{ij} y_{ij} = \max_{\mathfrak{X}} \sum a_{ij} x_{ij},$$

and moreover that the max is achieved only for elements of  $\mathfrak{X}$ .

► **LEMMA 5.8.2.** If  $\|y_{ij}^0\| \in \mathcal{Y}$  maximizes  $\sum_{i,j} a_{ij} y_{ij}$ , then  $\|y_{ij}^0\| \in \mathfrak{X}$ .

*Proof.* Let  $\sum_j y_{ij}^0 = r_i$  and  $\sum_i y_{ij}^0 = s_j$ ; then  $\sum_i r_i = \sum_j s_j = K$ . If one of the  $r_i$  (say  $r_{i_0}$ ) is less than 1, then one of the  $s_j$  (say  $s_{j_0}$ ) is less than 1, and conversely; and  $K$  is less than  $m$ .

Let

$$t_{ij} = \frac{(1 - r_i)(1 - s_j)}{m - K};$$

then obviously  $t_{i_0 j_0} > 0$ . Also  $\sum_j t_{ij} = 1 - r_i$  and  $\sum_i t_{ij} = 1 - s_j$ , so that  $t_{ij} + y_{ij}^0 \geq 0$  and  $\sum_j (t_{ij} + y_{ij}^0) = 1 - r_i + r_i = 1$  for all  $i, j$ . Similarly, the sum over  $i$  equals 1. Therefore  $\|t_{ij} + y_{ij}^0\| \in \mathfrak{X} \subset \mathcal{Y}$ .

But  $\sum_{i,j} a_{ij} (t_{ij} + y_{ij}^0) = \sum_{i,j} a_{ij} y_{ij}^0 + c$ , where  $c$  is a positive number and thus  $\|y_{ij}^0\|$  is not a maximum point. This contradicts the hypothesis, hence  $\|y_{ij}^0\| \in \mathfrak{X}$ .

The problem now has a linear programming form. For  $\mathbf{y}$  in  $m^2$ -space, we wish to maximize  $\sum_{i,j} a_{ij} y_{ij}$  subject to the constraints  $\mathbf{y} \geq \mathbf{0}$  and  $\mathbf{A}\mathbf{y} \leq \mathbf{u}$ , where  $\mathbf{A}$  is the matrix of Lemma 5.8.1 and  $\mathbf{u} = \langle 1, 1, \dots, 1 \rangle$  with  $2m$  components. Thus a solution of this linear programming problem leads to a solution of the optimal assignment problem.

The dual to the linear programming problem is the following. For  $\mathbf{z} = \langle t_1, \dots, t_m, s_1, \dots, s_m \rangle$ , minimize  $\sum t_i + \sum s_i$  subject to the constraints  $\mathbf{z} \geq \mathbf{0}$  and  $\mathbf{z}\mathbf{A} \geq \|a_{ij}\|$ . Written out in full, the constraints are

$$t_i + s_j \geq a_{ij}, \quad t_i \geq 0, \quad s_j \geq 0 \quad \text{for all } i, j.$$

(2) Von Neumann reduced the optimal assignment problem to the two-person zero-sum game with  $2m \times m^2$  matrix

$$\mathbf{B} = \begin{array}{c|ccccc} & & \overbrace{1/a_{11} \cdots 1/a_{1m}}^m & & & \\ m & \left\{ \begin{array}{c|ccccc} & & 1/a_{21} \cdots 1/a_{2m} & & & \\ & & \cdot & & & \\ & & & & & \\ & & & & & 1/a_{m1} \cdots 1/a_{mm} \\ & & & & & \end{array} \right. & & & \\ & \left\{ \begin{array}{c|ccccc} & 1/a_{11} & 1/a_{21} & \cdots & 1/a_{m1} & \\ & \cdot & \cdot & & \cdot & \\ & & & & & \\ & & 1/a_{1m} & 1/a_{2m} \cdots & & 1/a_{mm} \end{array} \right. & & & \\ m & \left\{ \begin{array}{c|ccccc} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{array} \right. & & & \\ m & \left\{ \begin{array}{c|ccccc} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{array} \right. & & & \end{array} \end{array}$$

with a zero entered in all places which are not obviously non zero. A solution to the assignment problem can be obtained from an optimal strategy for Player II in this game. Let

$$\mathbf{z}^0 = \langle z_{11}^0, \dots, z_{1m}^0, z_{21}^0, \dots, z_{2m}^0, \dots, z_{m1}^0, \dots, z_{mm}^0 \rangle$$

be an optimal strategy for Player II. Then the inner product of  $\mathbf{z}^0$  with the rows of  $\mathbf{B}$  must all be less than or equal to  $v$ , where  $v$  is the value of the game; that is,  $\mathbf{Bz}^0 \leq \mathbf{v}$ , where  $\mathbf{v} = \langle v, v, \dots, v \rangle$ . Since  $a_{ij} > 0$ , it follows that  $v > 0$ .

Let

$$y_{ij}^0 = \frac{z_{ij}^0}{a_{ij}v}.$$

Clearly,  $y_{ij}^0 \geq 0$ , and since  $\mathbf{z}^0$  is optimal, we obtain  $\sum_j y_{ij}^0 \leq 1$  for all  $i$  and  $\sum_i y_{ij}^0 \leq 1$  for all  $j$ . We wish to establish that  $\mathbf{y}^0$  maximizes  $\sum a_{ij}y_{ij}$ . Observe first that  $\sum_{ij} a_{ij}y_{ij}^0 = M = 1/v$ . Suppose  $\bar{\mathbf{y}} \in \mathcal{Y}$  and  $\sum_{ij} a_{ij}\bar{y}_{ij} = M' > M$ . Define

$$\bar{z}_{ij} = \frac{a_{ij}\bar{y}_{ij}}{M'}.$$

Then  $\bar{z}_{ij} \geq 0$  and  $\sum_{i,j} \bar{z}_{ij} = 1$ , so that  $\bar{\mathbf{z}}$  is a strategy for Player II in the game with matrix  $\mathbf{B}$ . The inner product of  $\bar{\mathbf{z}}$  with the  $i$ th row is

$$\sum_j \bar{z}_{ij} \frac{1}{a_{ij}} = \frac{1}{M'} \sum_j \bar{y}_{ij} \leq \frac{1}{M'} < v.$$

For the last  $m$  rows, we get

$$\sum_i \bar{z}_{ij} \frac{1}{a_{ij}} = \frac{1}{M'} \sum_i \bar{y}_{ij} \leq \frac{1}{M'} < v,$$

which contradicts the fact that  $v$  is the value of the game. Hence  $\mathbf{y}^0$  maximizes  $\sum a_{ij}y_{ij}$  for  $\mathbf{y} \in \mathfrak{X}$ .

But previously we found that if  $\mathbf{y}^0$  maximizes  $\sum a_{ij}y_{ij}$ , then  $\mathbf{y}^0 \in \mathfrak{X}$ ; consequently, a solution to the optimal assignment problem can be extracted from a representation of  $\mathbf{y}^0$  as a convex combination of extreme points.

An interpretation of the matrix game  $B$  as a type of hide-and-seek game can be given. Consider the matrix

$$\begin{vmatrix} 1/a_{11} & 1/a_{12} & \cdots & 1/a_{1m} \\ 1/a_{21} & 1/a_{22} & \cdots & 1/a_{2m} \\ \vdots & & & \\ 1/a_{m1} & 1/a_{m2} & \cdots & 1/a_{mm} \end{vmatrix}$$

Player I selects either a row or a column in this matrix and Player II selects a cell. If the cell is located in the row or column that Player I has chosen, Player I gets paid the value of that cell. Otherwise he receives nothing. The reader can verify that the matrix of this game is the matrix  $\mathbf{B}$ .

**5.9 Transportation and flow problem.** The transportation model was introduced in Section 5.1; the same notation will be used in this section. Stripped of economic interpretation, the transportation problem reads as follows: Minimize

$$\sum_{j=1}^n \sum_{i=1}^m x_{ij} c_{ij}$$

subject to the constraints

$$\begin{aligned} \sum_{i=1}^m x_{ij} &\geq r_j \quad (j = 1, \dots, n), \\ \sum_{j=1}^n x_{ij} &\leq s_i \quad (i = 1, \dots, m), \quad x_{ij} \geq 0 \quad \text{for all } i, j. \end{aligned} \tag{5.9.1}$$

By assumption,  $c_{ij} > 0$ ,  $r_j > 0$ , and  $s_i > 0$ . The existence of a solution is obvious.

Throughout the remainder of this section we shall first assume that  $\sum_j r_j \leq \sum_i s_i$  (i.e., that the supply is sufficient to satisfy demand).

► **LEMMA 5.9.1.** The solution to (5.9.1) is achieved only for vectors  $\|x_{ij}\|$  satisfying the relations

$$\sum_i x_{ij} = r_j, \quad \sum_j x_{ij} \leq s_i, \quad \text{and} \quad x_{ij} \geq 0.$$

*Proof.* Suppose that  $\mathbf{x}^0 = \|x_{ij}^0\|$  is a solution to (5.9.1), and that  $\sum_i x_{ij_0}^0 = b_{j_0} > r_{j_0}$  for some  $j_0$ . Then there exists  $i_0$  such that  $x_{i_0 j_0}^0 > 0$ . Set  $y_{ij} = x_{ij}^0$ , except that  $y_{i_0 j_0} = x_{i_0 j_0}^0 - \epsilon$ ; then if  $\epsilon > 0$  is chosen sufficiently small, it follows that

$$y_{ij} \geq 0, \quad \sum_i y_{ij} \geq r_j, \quad \text{and} \quad \sum_j y_{ij} \leq s_i.$$

Clearly,

$$\sum_{i,j} y_{ij} c_{ij} < \sum_{i,j} x_{ij}^0 c_{ij},$$

contradicting the fact that  $\mathbf{x}^0$  is a solution. Consequently,  $\sum_i x_{i,j}^0 = r_j$  for every  $j$ .

Henceforth, we impose the restriction that  $\sum_j r_j = \sum_i s_i$  (i.e., that the supply equals the demand). Then trivially every vector satisfying (5.9.1) necessarily obeys the equalities

$$\sum_i x_{ij} = r_j, \quad \sum_i x_{ij} = s_i \quad \text{and} \quad x_{ij} \geq 0. \quad (5.9.2)$$

Let  $\Gamma$  denote the collection of vectors  $\mathbf{x} = \|x_{ij}\|$  satisfying (5.9.2). From here on, in seeking a solution to (5.9.1), we consider only vectors belonging to  $\Gamma$ .

The reader may observe that the optimal assignment problem can be viewed as a special case of the transportation problem, in which  $r_i = s_j = 1$  and  $m = n$ . It is therefore reasonable to expect many of the results valid for the assignment problem to have analogs for the transportation problem. Some of these will be pointed out presently.

The fundamental constraint matrix of the transportation problem is similar to (5.8.2):

$$\mathbf{A} = \left( \begin{array}{cccccc} x_{11}, \dots, x_{1n}, & x_{21}, \dots, x_{2n}, & \dots & x_{m1}, \dots, x_{mn} \\ \left. \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right\} m & \left. \begin{array}{c} \dots \\ \vdots \\ \dots \\ 1 \end{array} \right\} n & & & & \left. \begin{array}{c} \dots \\ \vdots \\ \dots \\ 1 \end{array} \right\} n \\ 1 & 1 & \dots & 1 & \dots & 1 \\ \left. \begin{array}{c} \dots \\ \vdots \\ \dots \\ 1 \end{array} \right\} n & \left. \begin{array}{c} \dots \\ \vdots \\ \dots \\ 1 \end{array} \right\} n & & & & \left. \begin{array}{c} \dots \\ \vdots \\ \dots \\ 1 \end{array} \right\} n \end{array} \right)$$

Let  $\omega = \langle s_1, \dots, s_m, r_1, \dots, r_n \rangle$ . Then the constraint set of (5.9.2) may be expressed by the vector equalities  $\mathbf{Ax} = \omega$ .

In studying any linear programming problem it is desirable to characterize the extreme points of the constraint set, since the optimum must occur at one of the extreme points. For this purpose the following lemma is useful.

► **LEMMA 5.9.2.** If  $r_i$  and  $s_j$  not all zero are all integral (i.e., integer-valued), then the extreme points of  $\Gamma$  are contained in the set of all  $\mathbf{x} = \|x_{ij}\|$  of  $\Gamma$  with  $x_{ij}$  integral. Moreover, there exists at least one row or column with at most one nonzero entry.

The proof is based on the same ideas as that of Lemma 5.8.1. The arguments of Lemma 5.8.1 show that any extreme point  $\mathbf{x} \in \Gamma$  is composed of at most  $m + n - 1$  nonzero components. Consequently, there exists a row or a column, say a row, with at most one nonzero element, which must be integral (since the  $r_i$ 's are integral). By removing that row and changing one of the  $s_j$ 's appropriately, we arrive at an extreme point  $\mathbf{x}$  involving  $m - 1$  rows and at most  $n$  columns, with the new  $r_i$ 's and  $s_j$ 's still integral. Repeated applications of this reasoning prove the lemma.

It is an easy step from Lemma 5.9.2 to the following theorem.

► **THEOREM 5.9.1.** If  $r_i$  and  $s_j$  are integral, then there exists a solution  $\mathbf{x}^0$  to (5.9.1) with integral components  $x_{ij}^0$ .

Theorem 5.9.1 is very appealing. It says that if supply and demand are expressed in terms of integral, indivisible quantities of a commodity—e.g., 50 refrigerators, 2000 sacks of salt—then the optimal strategy for minimizing costs requires shipping only integral amounts of the commodity from the plants to the various markets, even though *a priori* we permitted strategies involving the transfer of fractional amounts of the commodity.

*Network flow problems.* The remainder of this section is devoted to the study of network flow problems. This model is closely related, and in certain respects equivalent, to the transportation model. The network problem arises naturally in the study of transportation networks. One is given a network of directed arcs and nodes with two distinguished nodes, one called the source and the other the sink.\* A directed arc may be

\* The problem in which there are several sources and sinks with flows permitted from any source to any sink is reducible to a single-source, single-sink problem.

The case where flow is allowed in either direction with a restriction on the total capacity of each arc can be reduced to the previous problem with all arcs duplicated once each way. The restriction of the total is no more than the same restriction of flow in each direction, there being no advantage to having flow both ways at once (see also Problem 15 of Chapter 6).

The separate consideration of nodes may similarly be obviated by introducing fictitious arcs from each node to itself (see Problem 14 of Chapter 6).

interpreted as a railway connection between two terminals, and nodes may be thought of as warehouses located at the various intermediate stations. Each directed arc in the network has associated with it a nonnegative integer, its flow capacity; each node also has its capacity, which may be thought of as the capacity of a warehouse. Source arcs (arcs having a terminus at the source) may be assumed to be directed away from the source; sink arcs are always directed toward the sink.

It is desired to find a maximal flow from source to sink, subject to two conditions: (i) the flow in an arc must be in the direction of the arc and may not exceed its capacity, and the flow into a node may not exceed its capacity; (ii) the total flow into any intermediate node must be equal to the flow out of it. The value of any flow is defined to be the sum of the flows in the source arcs. Since any flow is consistent with the capacity restraints, the value of a flow is also equal to the sum of flows in the sink arcs. For example, consider Fig. 5.1, where source and sink are denoted by  $+$  and  $-$ , respectively, and the capacity limitations are as indicated. The maximal flow for this network is described in Fig. 5.2.

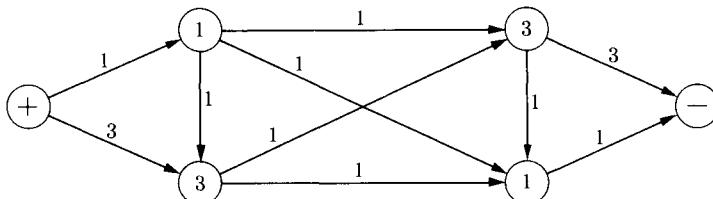


FIGURE 5.1

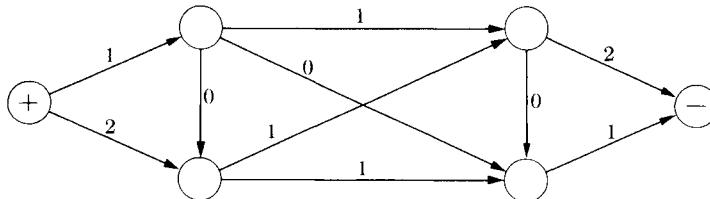


FIGURE 5.2

Let  $x_{ij} \geq 0$  ( $i \neq j; i, j = 1, \dots, m$ ) denote the flow from node  $i$  to node  $j$ ;  $x_{ii}$  ( $i = 1, \dots, m$ ) the total flow through node  $i$ ;  $x_{0j}$  ( $j = 1, \dots, m$ ) the flow from the source to node  $j$ ; and  $x_{i0}$  ( $i = 1, \dots, m$ ) the flow from node  $i$  to the sink.

For the general network flow problem, we consider the pseudo-transportation array

$$\begin{array}{ccccc}
 & & & & \text{Totals} \\
 \left[ \begin{array}{ccccc}
 -x_{00} & x_{01} & x_{02} & \cdots & x_{0m} \\
 x_{10} & -x_{11} & x_{12} & \cdots & x_{1m} \\
 \vdots & & & & \vdots \\
 x_{m0} & x_{m1} & x_{m2} & \cdots & -x_{mm}
 \end{array} \right] & & & & 0 \\
 \text{Totals} & 0 & 0 & 0 & \cdots & 0
 \end{array}$$

schematizing the equations

$$\begin{aligned}
 -x_{ii} + \sum_{\substack{j \\ j \neq i}} x_{ij} &= 0 \quad (i = 0, 1, \dots, m), \\
 -x_{jj} + \sum_{\substack{i \\ i \neq j}} x_{ij} &= 0 \quad (j = 0, 1, \dots, m).
 \end{aligned} \tag{5.9.3}$$

Clearly,  $x_{00}$  is the total flow, and the problem is to maximize  $x_{00}$  subject to (5.9.3) and

$$0 \leq x_{ij} \leq a_{ij} \quad (i, j) \neq (0, 0), \tag{5.9.4}$$

where the  $a_{ij}$  are the given nonnegative constants that describe the capacities of the nodes and arcs. Arcs not present in the network have  $a_{ij} = 0$ .

This is not a transportation problem in the strict sense because of the minus signs down the diagonal in the array. Notice that the constraining equations (5.9.3) are similar to relations (5.9.2) with all  $r_j = s_i = 0$ . However, transportation theory (e.g., Theorem 5.9.1) carries over directly to problems of this kind, even with upper bound constraints present. We omit the simple proof of this fact.

**\*5.10 Maximal-flow, minimal-cut theorem.** A chain is any collection of directed arcs and nodes which leads from source to sink. A “cut” in a directed network is a collection of directed arcs and nodes meeting every directed chain from source to sink. The value of a cut is the sum of the capacities of all its member nodes and arcs. It is clear that the maximal flow cannot exceed the minimal cut value, and conversely. Hence we have the following theorem.

► **THEOREM 5.10.1.** The maximal flow value is equal to the minimal cut value.

This theorem is essentially a rephrasing of the duality theorem (Theorem 5.4.1) for the network flow problem. In order to provide a proof, we shall examine the form of the dual problem.

Let

$$\mathbf{A} = \begin{array}{c|c} & \begin{array}{cccccc} x_{00}, x_{01}, \dots, x_{0m}, & x_{10}, x_{11}, \dots, x_{1m}, & \dots, & x_{m0}, x_{m1}, \dots, x_{mm} \end{array} \\ \left. \begin{array}{c} m+1 \\ \vdots \\ m+1 \\ \vdots \\ m^2+2m \end{array} \right\} & \left[ \begin{array}{cccccc} -1 & 1 & \cdots & 1 & & & \\ & 1 & -1 & \cdots & 1 & & \\ & & & & \ddots & & \\ & & & & & 1 & 1 & \cdots & -1 \\ -1 & & & 1 & & \cdots & 1 & 1 & \\ & 1 & & -1 & & & & 1 & \\ & & \ddots & & \ddots & & & \ddots & \\ & & & 1 & & 1 & \cdots & & -1 \\ 0 & & & & & I & & & \\ \vdots & & & & & & & & \\ 0 & & & & & & & & \end{array} \right] \end{array},$$

where  $I$  stands for the identity matrix of size  $m^2 + 2m$ .

Let

$$\mathbf{b} = \underbrace{\langle 0, 0, \dots, 0 \rangle}_{2m+2}, \langle a_{01}, a_{02}, \dots, a_{0m}, a_{10}, a_{11}, \dots, a_{mm} \rangle;$$

then the matrix of constraints becomes  $\mathbf{Ax} \leq \mathbf{b}$ , with equality for the first  $2m + 2$  components. Since the flow problem seeks to maximize  $x_{00}$ , the system of dual constraints in the dual variables  $u_i$  ( $i = 0, \dots, m$ )  $v_j$  ( $j = 0, 1, \dots, m$ ), and  $\omega_{ij}$  [with  $(i, j) \neq (0, 0)$ ] becomes

$$\begin{aligned} u_i + v_j + \omega_{ij} &\geq 0 & (i, j = 0, \dots, m; i \neq j), \\ -u_i - v_i + \omega_{ii} &\geq 0 & (i = 1, \dots, m), \\ -u_0 - v_0 &\geq 1, \\ \omega_{ij} &\geq 0 & (i, j) \neq (0, 0). \end{aligned} \tag{5.10.1}$$

Let the set of all  $(u, v, \omega)$  satisfying (5.10.1) be denoted by  $T$ . Since the original constraints involved  $2m + 2$  equalities, the variables  $u_i$  and  $v_j$  vary unrestrictedly (see page 117). The optimal  $\mathbf{x}$  strategy (except in trivial cases) must have  $x_{00} > 0$  by Theorem 5.4.3; hence  $-u_0 - v_0 = 1$ . Moreover, because one of the equations (5.9.3) is redundant, we may arbitrarily select  $u_0 = 0$ , which implies that  $v_0 = -1$ .

At this point the following lemma is needed.

►LEMMA 5.10.1. If  $u_0 = 0$  and  $v_0 = -1$ , the extreme points of  $T$  have  $u_i$  equal to 0 or 1 for all  $i$ ,  $v_j$  equal to 0 or  $-1$  for all  $j$ , and  $\omega_{ij}$  equal to 0 or 1 for all  $i, j$ .

*Proof.* The system (5.10.1) comprises  $m^2 + 4m$  variables and  $2m^2 + 4m$  inequalities. The extreme points of  $T$  are obtained by finding the solutions to all possible sets of  $m^2 + 4m$  independent equations in the same number of unknowns. The total set of equations obtainable from (5.10.1) reduces to

$$\begin{aligned} v_j + \omega_{0j} &= 0, \quad u_i + \omega_{i0} = 1, \\ u_i + v_j + \omega_{ij} &= 0 \quad (i, j = 1, \dots, m; i \neq j), \\ \omega_{ij} &= 0 \quad (i, j \neq 0, 0), \\ -u_i - v_i + \omega_{ii} &= 0 \quad (i = 1, \dots, m). \end{aligned} \quad (5.10.2)$$

Substituting  $\omega_{ij} = 0$  wherever this is so, we have a system of equations in the remaining  $\omega_{ij}$ ,  $u_i$ , and  $v_j$ .

If in the system (5.10.2) the equation  $\omega_{i_0j_0} = 0$  is not present, then the associated variable  $\omega_{i_0j_0}$  has to be determined by one of the remaining equations which is necessarily of the form

$$\begin{aligned} u_i + v_j + \omega_{ij} &= 0 \quad (i, j \neq 0, 0 \text{ and } i \neq j), \\ -u_i - v_i + \omega_{ii} &= 0 \quad (i \neq 0), \\ v_j + \omega_{0j} &= 0 \quad (j \neq 0), \end{aligned}$$

or

$$u_i + \omega_{i0} = 1 \quad (i \neq 0).$$

These equations cannot determine  $u_i$  and  $v_j$ . But we see that  $u_i$  and  $v_j$  with  $i, j$  not zero are determined by  $2m$  independent equations, each of which has the form

$$u_i + v_j = 0, \quad v_j = 0, \quad \text{or} \quad u_i = 1. \quad (5.10.3)$$

Each  $u_i$  ( $i = 1, \dots, m$ ) and  $v_j$  ( $j = 1, \dots, m$ ) is included in some equation of (5.10.3). It is readily seen that a system of independent equations of the type (5.10.3) involving all the variables  $u_i$  and  $v_j$  ( $i, j = 1, \dots, m$ ) can only have solutions with  $u_i$  equal to 0 or 1 and  $v_j$  equal to 0 or  $-1$ . On examination of (5.10.2), since  $\omega_{ij} \geq 0$ , it then follows that  $\omega_{ij}$  equals 0 or 1 as asserted.

*Proof of Theorem 5.10.1.* The dual programming problem reduces to the form of minimizing  $\sum_{ij} a_{ij} \omega_{ij}$  subject to (5.10.1). Because of the duality theorem (Theorem 5.4.1) and Lemma 5.10.1,

$$\max_{\sigma} x_{00} = \sum_{\sigma} a_{ij}, \quad (5.10.4)$$

where  $\sigma$  is the set of arcs  $i, j$  and nodes  $i$  corresponding to  $\omega_{ij} = 1$ . By Theorem 5.4.3,  $\omega_{ij} = 1$  implies  $x_{ij} = a_{ij}$  and  $\omega_{ii} = 1$  implies  $x_{ii} = a_{ii}$ .

It is asserted that  $\sigma$  constitutes a cut. For suppose that all  $a_{ij}$  such that  $i, j \notin \sigma$  are increased by  $\epsilon > 0$ . That this increase does not change the solution to the dual problem is evident from (5.10.4) and hence it cannot increase the flow in the network. But if there were some directed chain from source to sink not meeting  $\sigma$ , clearly the flow could be increased. Thus  $\sigma$  is a cut, and since no flow can exceed the value of any cut,  $\sigma$  is a minimal cut.

**\*5.11 The caterer's problem.** In order to serve meals during the next  $m$  days, a caterer will need  $r_j$  fresh napkins on the  $j$ th day. Normal laundering takes  $p$  days at  $b$  cents a napkin; rapid laundering takes  $q$  days ( $p$  and  $q$  are integers) at  $c$  cents a napkin. Also, new napkins may be purchased at  $a$  cents a napkin. How can the caterer arrange matters to meet his needs and minimize his outlays for the  $m$  days?

Another interpretation of the model involves military aircraft. "Laundering" corresponds to overhaul of engines, the unit of time is a month rather than a day, etc.

Practical considerations dictate that  $r_j \geq 0$  ( $j = 1, \dots, m$ ),  $a > b$ ,  $a > c$ ,  $c > b$ ,  $p > q$ . Let  $x_j$  be the number of napkins purchased for use on the  $j$ th day,  $y_j$  the number sent for normal laundry service,  $z_j$  the number sent for rapid laundry service, and  $s_j$  the number of soiled napkins on hand on the  $j$ th day ( $j = 1, \dots, m$ ). Also, let capital letters denote cumulative amounts, e.g.,

$$X_j = \sum_{i=1}^j x_i, \quad \text{etc.}$$

The caterer's problem is as follows.

*Problem A.* Minimize the total cost

$$a \sum_{j=1}^m x_i + b \sum_{j=1}^m y_j + c \sum_{j=1}^m z_j = aX_m + bY_m + cZ_m \quad (5.11.1)$$

subject to the constraints

$$x_j + y_{j-p} + z_{j-q} \geq r_j, \quad (5.11.2)$$

$$y_j + z_j + s_j - s_{j-1} = r_j, \quad (5.11.3)$$

$$x_j, y_j, z_j, s_j \geq 0, \quad (5.11.4)$$

where  $j = 1, \dots, m$ ,  $s_0 = 0$ , and  $y_{-k} = z_{-k} = 0$  for  $k \geq 0$ .

Relation (5.11.2) expresses the need requirement and (5.11.3) states what happens. Implicit in the formulation of these constraints is the reasonable assumption that no napkin will be bought or laundered ahead of time.

It is possible to formulate the caterer's problem as a distribution or transportation problem in which the caterer and each day's hamper of soiled napkins are the origins, and each day's requirement of fresh napkins and the final inventory of soiled napkins are the destinations (see Problem 10 of Chapter 6). Consequently, some of the special computational schemes for the transportation problem can be advantageously adapted to provide a numerical solution to the caterer's problem.

For our purposes here, however, certain features of the caterer's problem are more important than its numerical solution. In the ensuing analysis we follow the methods of Jacobs, who obtained an explicit analytical solution for  $p - 1 = q$ . Jacobs' approach elegantly displays the technique of reducing a linear programming problem of complex form to a problem of simpler form.

As regards Problem A, two simplifications can be made. First, it is clear that restrictions (5.11.2) may be replaced by

$$x_j + y_{j-p} + z_{j-q} = r_j \quad (j = 1, \dots, m). \quad (5.11.2a)$$

Second, napkins are not laundered unless they can be returned by the  $m$ th day; so let  $y_j = 0$  ( $m - p < j \leq m$ ) and  $z_j = 0$  ( $m - q < j \leq m$ ). Then the cost function becomes

$$aX_m + bY_{m-p} + cZ_{m-q}.$$

Noticing that  $X_m + Y_{m-p} + Z_{m-q} = R_m$ , which can be obtained by summing (5.11.2a) over  $j = 1, \dots, m$ , we get

$$\begin{aligned} aX_m + bY_{m-p} + cZ_{m-q} &= (a - b)X_m + (c - b)Z_{m-q} + bR_m \\ &= (c - b) \left\{ \frac{a - b}{c - b} X_m + Z_{m-q} \right\} + bR_m. \end{aligned}$$

Now Problem A is reduced to

*Problem B.* Minimize

$$\lambda X_m + Z_{m-q} \quad (5.11.5)$$

subject to the constraints (5.11.2a), (5.11.3), and (5.11.4), where

$$\lambda = \frac{a - b}{c - b} > 0.$$

Now we shall transform the variables in order to eliminate the role of the individual  $x_j$ . Accumulating (5.11.2a) up to  $j$ , we obtain

$$X_j + Y_{j-p} + Z_{j-q} = R_j; \quad (5.11.6)$$

and accumulating (5.11.3) up to  $j = p$ , we obtain

$$Y_{j-p} + Z_{j-p} + s_{j-p} = R_{j-p}. \quad (5.11.7)$$

Subtracting (5.11.7) from (5.11.6), we obtain

$$X_j + Z_{j-q} - Z_{j-p} - s_{j-p} = R_j - R_{j-p}. \quad (5.11.8)$$

Let the transformation of  $(x, y, s, z)$  into  $(w, u, v, z)$  be defined as follows:

$$w = X_m, \quad u_j = X_m - X_j + s_{j-p}, \quad v_j = x_j + y_{j-p}, \quad z_j = z_j.$$

Then (5.11.8), (5.11.2a), and (5.11.4) will be written

$$w + Z_{j-q} - Z_{j-p} - u_j = R_j - R_{j-p}, \quad (5.11.9)$$

$$z_{j-q} + v_j = r_j, \quad (5.11.10)$$

$$w, z_j, u_j, v_j \geq 0, \quad (5.11.11)$$

and the function (5.11.5) to be minimized will be

$$\lambda w + Z_{m-q}. \quad (5.11.12)$$

We shall now prove

► **LEMMA 5.11.1.** Problem B can be reduced to

*Problem C.* Minimize (5.11.12) subject to the constraints (5.11.9), (5.11.10), and (5.11.11).

*Proof.* As noted above, to each feasible  $(x, y, s, z)$  of Problem B corresponds a feasible  $(w, u, v, z)$  of Problem C. The converse is not true. However, every solution of Problem C is the image of a feasible  $(x, y, s, z)$  of Problem B.

Let  $(w, u, v, z)$  be optimal and define  $X_m = w$  and

$$X_j = w - \min_{i \leq j} u_i$$

so that  $X_j$  is nondecreasing. The first relation of (5.11.9) implies that  $w \geq u_1$ ; hence the  $x_j = X_j - X_{j-1}$  are all nonnegative. Now

$$\min_{1 \leq j \leq m} u_j = 0;$$

for otherwise  $w$  and every  $u_j$  could be decreased by a fixed amount, contradicting the optimality assumption. Consequently, the two definitions of  $X_m$  coincide. Furthermore,  $X_m - X_j \leq u_j$ ; hence  $s_{j-p} = u_j - X_m + X_j \geq 0$ .

It remains to verify that the  $y_j = v_{j+p} - x_{j+p}$  are nonnegative. If  $x_k = 0$ , where  $k = j + p$ , then we are finished. So let us assume that  $x_k > 0$ . Then from the definition of  $x_k$ , it follows that  $X_k = X_m - u_k$  and  $u_i > 0$  for  $i < k$ . Suppose  $z_{k-p}$  positive; then if we reduce  $z_{k-p}$  a fixed amount and each of  $u_{k-p+q}, u_{k-p+q+1}, \dots, u_{k-1}$  by the same amount and change  $v_{k-p+q}$  appropriately, the constraints of Problem C remain unaffected while the  $Z_{m-q}$  are decreased. This contradicts the optimality assumption; hence  $z_{k-p} = 0$ . Now  $s_{k-p} = u_k - X_m + X_k = 0$ ; therefore

$$y_j = y_{k-p} = r_{k-p} + s_{k-p-1} - s_{k-p} - z_{k-p} = 0 + r_{k-p} + s_{k-p-1} \geq 0.$$

The proof of the lemma is complete.

In view of Lemma 5.11.1, the problem can now be stated as follows.

*Problem C'.* Minimize  $[\lambda w + Z_{m-q}]$  subject to the constraints

$$w + Z_{j-q} - Z_{j-p} \geq R_j - R_{j-p} \quad (j = 1, 2, \dots, m), \quad (5.11.13)$$

$$0 \leq z_{j-q} \leq r_j \quad (j = 1, 2, \dots, m), \quad (5.11.14)$$

$$0 \leq w. \quad (5.11.15)$$

(The reader may note that we have converted the constraints of Problem C into an equivalent form involving inequalities.)

An algorithm for solutions of the caterer's problem with arbitrary  $p$  and  $q$  is summarized by Problems 22–26 (see also the notes to this section at the close of the chapter). Here we shall limit our attention to the case  $q = p - 1$  and give an explicit form for a solution of Problem C'.

For  $q = p - 1$ , (5.11.13) reduces to

$$w + z_{j-q} \geq R_j - R_{j-q-1} \quad (j = 1, \dots, m). \quad (5.11.16)$$

From (5.11.15) and (5.11.14), we have

$$w \geq R_j - R_{j-q} \quad (j = 1, \dots, m). \quad (5.11.17)$$

Let

$$\bar{w} = w - \max_{1 \leq i \leq m} \{R_{i-1} - R_{i-p}\},$$

$$H_j = R_j - R_{j-p} - \max_{1 \leq i \leq m} \{R_{i-1} - R_{i-p}\}.$$

Then (5.11.16) and (5.11.17) can be written as

$$\bar{w} + z_{j-q} \geq H_j, \quad \bar{w} \geq 0.$$

It will now be shown that Problem C' is reduced to the following form.

*Problem C''.* Minimize  $\lambda\bar{w} + Z_{m-q}$  subject to the constraints

$$\bar{w} + z_{j-q} \geq H_j, \quad \bar{w} \geq 0, \quad z_{j-q} \geq 0.$$

*Proof.* It is clear that a solution to Problem C' should be feasible for Problem C''.

Let  $\bar{w}, z_j$  be a solution to Problem C''. Then we shall show that (5.11.14) is indeed satisfied. Since

$$H_j \leq R_j - R_{j-p} - (R_{j-1} - R_{j-p}) = r_j,$$

any  $z_{j-q} > r_j$  would not be optimal.

The solution to Problem C'' is implied by the following lemma.

►LEMMA 5.11.2. Let  $a_1 \geq \dots \geq a_m \geq 0$ . A solution to the problem of minimizing  $\lambda w + Z_m$  subject to the constraints  $w + z_j \geq a_j$ ,  $w \geq 0$ ,  $z_j \geq 0$  ( $j = 1, \dots, m$ ) is given by

$$w = a_{[\lambda]+1}, z_j = \max \{0, a_j - a_{[\lambda]+1}\}$$

for  $\lambda < m$ . If  $\lambda \geq m$ , the minimum solution is given by

$$w = 0, \quad z_j = a_j.$$

*Proof.* For any given  $w \geq 0$ ,  $Z_m$  is minimized when

$$z_j = \max \{0, a_j - w\} \quad (j = 1, \dots, m).$$

Let  $a_k \geq w_k \geq a_{k+1}$ , and consider the function

$$\begin{aligned} \phi(w_k) &= \lambda w_k + \sum_{j=1}^m \max \{0, a_j - w_k\} = \lambda w_k + \sum_{j=1}^k (a_j - w_k) \\ &= (\lambda - k)w_k + \sum_{j=1}^k a_j; \end{aligned}$$

then

$$\begin{aligned} \phi(w_k) - \phi(w_{k+1}) &= (\lambda - k)(w_k - w_{k+1}) - (a_{k+1} - w_{k+1}) \\ &\leq 0 \quad \text{if} \quad k \geq \lambda \\ &\geq 0 \quad \text{if} \quad k < \lambda - 1. \end{aligned}$$

Therefore  $\phi(w_k)$  will be minimized when

$$w_k = a_k, \quad k = [\lambda] + 1,$$

as asserted.

Translating the final result to the initial problem, we obtain the solution of Problem A as

$$\begin{aligned} z_j &= \max \{(H_{j+p-1} - \bar{w}), 0\}, \\ x_j &= \max_{i \leq j} \{R_i - R_{i-p} - z_{i-p+1}\} = \max_{i \leq j-1} \{R_i - R_{i-p} - z_{i-p+1}\}, \\ y_j &= r_{j+p} - x_{j+p} - z_{j+1} \quad [\text{from (5.11.2a)}], \\ s_j &= R_j - Y_j - Z_j \quad [\text{from (5.11.7)}], \end{aligned}$$

where

$$H_j = \max \{(R_j - R_{j-p} - \max_{1 \leq i \leq m} \{R_i - R_{i-p}\}), 0\}$$

and  $\bar{w} = H_{[\lambda]+1}$ , with  $[\lambda]$  the largest integer smaller than or equal to  $(a-b)/(c-d)$ .

**\*5.12 Price speculation model.** Various securities can be bought or sold in the course of  $m + 1$  periods of time. The prices  $b_k$  during the  $k$ th period are predictable and hence supposedly known. In spite of this unrealistic assumption, the analysis of this deterministic problem is of some interest.

Let  $X_{k-1}$  denote the amount of securities on hand at the beginning of the  $k$ th period. During that period the change in value is  $b_k(X_{k-1} - X_k)$ ; if  $X_k > X_{k-1}$ , stock has been purchased, while if  $X_k < X_{k-1}$ , stock has been sold. The handling costs of the stock during a period are taken to be proportional to the amount of stock on hand; but these costs are not paid until the  $m$  periods are over, and thus they do not affect the amount of money available for speculation. The net receipts for the  $k$ th period are therefore

$$b_k(X_{k-1} - X_k) - cX_k.$$

Taking account of the interest rate  $i$ , the total receipts for all  $m + 1$  periods are

$$\sum_{k=0}^m \alpha^k [b_k(X_{k-1} - X_k) - cX_k], \quad \text{where } \alpha = \frac{1}{1+i} \quad (5.12.1)$$

with  $X_{-1} = 0$ . The amount of money  $M_k$  on hand at the end of the  $k$ th period satisfies the recursion relation

$$M_k = M_{k-1} + b_k(X_{k-1} - X_k), \quad (5.12.2)$$

with  $M_0$ , the initial amount of money available, fixed. It is desired to determine the speculation policy  $X_k$  which maximizes the total receipts subject to the constraint  $M_k \geq 0$  for all  $k$ . In other words, the speculator never owes money, and borrowing is not allowed.

The constraint inequalities can be expressed as

$$0 \leq X_k \leq \frac{M_0 + \sum_{i=0}^{k-1} (b_{i+1} - b_i)X_i}{b_k} \quad (k = 1, \dots, m), \quad (5.12.3)$$

where  $b_0 = 0$ . Subject to the conditions of (5.12.3) our problem is to determine the policy  $X$  which maximizes

$$\sum_{j=0}^{m-1} \alpha^j (\alpha b_{j+1} - b_j - c) X_j + \alpha^m X_m (-b_m - c). \quad (5.12.4)$$

Obviously, the maximum is achieved for  $X_m = 0$  (no stocks should be kept at the end of time  $m$ ).

The analysis proceeds by determining the optimal  $X_k$  backward in time. This is reasonable, since the constraint relation for  $X_i$  involves only the preceding  $X$ 's. For example, if the coefficient  $\alpha^{m-1}(\alpha b_m - b_{m-1} - c)$  of  $X_{m-1}$  is positive, then the optimal  $X_{m-1}$  in (5.12.3) must satisfy the right-hand equality

$$X_{m-1} = \frac{M_0 + \sum_{i=0}^{m-2} (b_{i+1} - b_i)X_i}{b_{m-1}}, \quad (5.12.5)$$

provided that this expression is positive.

We shall find that the positive values of an optimal policy  $X_i$  must correspond to a subset of the indices  $i$  with the property  $b_{i+1} - b_i > 0$ . Consequently, the upper bounds in (5.12.3) are always positive. In short, when prices are dropping, there is no advantage in maintaining stock. However, it still might not pay to keep stock when prices rise too slowly.

We now direct our attention to the explicit determination of the solution. For this purpose some useful implications of the constraint relation (5.12.5) should be noted. Observe that if (5.12.5) holds (that is,  $\alpha b_m - b_{m-1} - c > 0$ ), then the coefficient of  $X_{m-2}$  in (5.12.4) reduces to

$$\alpha^{m-2}(\alpha b_{m-1} - b_{m-2} - c) + \frac{b_{m-1} - b_{m-2}}{b_{m-1}} \alpha^{m-1}(\alpha b_m - b_{m-1} - c). \quad (5.12.6)$$

Hence, if (5.12.6) is positive, then  $X_{m-2}$  is to be chosen as large as possible so that [see (5.12.3)]

$$X_{m-2} = \frac{M_0 + \sum_{i=0}^{m-3} (b_{i+1} - b_i)X_i}{b_{m-2}}. \quad (5.12.7)$$

But upon substituting the value of  $X_{m-2}$  in (5.12.7) into (5.12.5), we obtain

$$X_{m-1} = \frac{M_0 + \sum_{i=0}^{m-3} (b_{i+1} - b_i)X_i}{b_{m-2}}.$$

This phenomenon constantly appears: whenever an  $X_i$  satisfies its constraint relation with equality in its upper bound in (5.12.3), then all the relations of the higher  $X_i$  can be reduced. Furthermore, if (5.12.6) is positive, it follows that  $b_{m-1} > b_{m-2}$ . If (5.12.6) is negative, then  $X_{m-2} = 0$  and (5.12.5) reduces to

$$X_{m-1} = \frac{b_{m-2}}{b_{m-1}} \frac{M_0 + \sum_{i=0}^{m-3} (b_{i+1} - b_i)X_i}{b_{m-2}}.$$

We now describe the general case. The construction is inductive. Assuming that relations for  $X_i$  with  $r_k + 1 \leq i \leq m - 1$  have been determined, we proceed to derive relations for  $X_{r_k}$ .

First, let us assume for definiteness that  $\alpha b_m - b_{m-1} - c > 0$ , so that  $X_{m-1} > 0$ . The reader can readily furnish the modifications in the contrary case.

*Case 1.* Let

$$X_i = 0 \quad \text{for } r_2 < i \leq r_1, \quad r_4 < i \leq r_3, \quad \dots, \quad r_k < i \leq r_{k-1},$$

$$\begin{aligned} X_i &= \frac{b_{r_2}}{b_{r_1}} \cdot \frac{b_{r_4}}{b_{r_3}} \cdots \frac{b_{r_k}}{b_{r_{k-1}}} \left( \frac{M_0 + \sum_{j=0}^{r_k} (b_{j+1} - b_j)X_j}{b_{r_k}} \right) && \text{for } r_1 < i \leq m - 1, \\ X_i &= \frac{b_{r_4}}{b_{r_3}} \cdots \frac{b_{r_k}}{b_{r_{k-1}}} \left( \frac{M_0 + \sum_{j=0}^{r_k} (b_{j+1} - b_j)X_j}{b_{r_k}} \right) && \text{for } r_3 < i \leq r_2, \\ &\vdots \\ X_i &= \frac{b_{r_k}}{b_{r_{k-1}}} \left( \frac{M_0 + \sum_{j=0}^{r_k} (b_{j+1} - b_j)X_j}{b_{r_k}} \right) && \text{for } r_{k-1} < i \leq r_{k-2}, \end{aligned} \quad (5.12.8)$$

with  $m - 1 > r_1 > r_2 > \dots > r_{k-1} > r_k$ . Inserting these expressions in (5.12.4), we see that the coefficient of  $X_{r_k}$  is

$$\begin{aligned} \alpha^{r_k} (\alpha b_{r_{k+1}} - b_{r_k} - c) &+ \frac{b_{r_{k+1}} - b_{r_k}}{b_{r_k}} \left\{ \frac{b_{r_k}}{b_{r_{k-1}}} \sum_{r_{k-1} < i \leq r_{k-2}} \alpha^i (\alpha b_{i+1} - b_i - c) \right. \\ &+ \frac{b_{r_{k-2}}}{b_{r_{k-3}}} \frac{b_{r_k}}{b_{r_{k-1}}} \sum_{r_{k-4} < i \leq r_{k-3}} \alpha^i (\alpha b_{i+1} - b_i - c) + \dots \\ &\left. + \frac{b_{r_2}}{b_{r_1}} \frac{b_{r_4}}{b_{r_3}} \cdots \frac{b_{r_k}}{b_{r_{k-1}}} \sum_{r_1 < i \leq m-1} \alpha^i (\alpha b_{i+1} - b_i - c) \right\}. \end{aligned} \quad (5.12.9)$$

Therefore, if (5.12.9) is negative, the optimal value of  $X_{r_k}$  must be 0. In this circumstance, we examine the coefficient of  $X_{r_{k-1}}$ , after first reducing all the relations (5.12.8) and observing that they are again of the same form. If (5.12.9) is positive, then  $b_{r_{k+1}} > b_{r_k}$  and  $X_{r_k}$  must equal its upper bound in (5.12.3). We are thus led to a second set of relations, which are valid when the last group of  $X$ 's determined has satisfied the constraint relation (5.12.3) with equality in the upper bound. These are of the following form.

*Case 2.*

$$X_i = 0 \quad \text{for } r_2 < i \leq r_1, \quad r_4 < i \leq r_3, \quad \dots, \quad r_k < i \leq r_{k-1} \\ \text{and } r_{k+1} < r_k,$$

$$X_i = \frac{b_{r_2} \cdot b_{r_4} \cdots b_{r_k}}{b_{r_1} \cdot b_{r_3} \cdots b_{r_{k-1}}} \frac{M_0 + \sum_{j=0}^{r_k+1} (b_{j+1} - b_j) X_j}{b_{r_{k+1}+1}} \\ \text{for } r_1 < i \leq m-1,$$

$$X_i = \frac{b_{r_4} \cdots b_{r_k}}{b_{r_3} \cdots b_{r_{k-1}}} \frac{M_0 + \sum_{j=0}^{r_k+1} (b_{j+1} - b_j) X_j}{b_{r_{k+1}+1}} \quad (5.12.10) \\ \text{for } r_3 < i \leq r_2,$$

⋮

$$X_i = \frac{b_{r_k}}{b_{r_{k-1}}} \frac{M_0 + \sum_{j=0}^{r_k+1} (b_{j+1} - b_j) X_j}{b_{r_{k+1}+1}} \\ \text{for } r_{k-1} < i \leq r_{k-2},$$

$$X_i = \frac{M_0 + \sum_{j=0}^{r_k+1} (b_{j+1} - b_j) X_j}{b_{r_{k+1}+1}} \\ \text{for } r_{k+1} < i \leq r_k.$$

The decision about  $X_{r_{k+1}}$  depends on the sign of its coefficient  $L_{r_{k+1}}$ , obtained by substituting the relations (5.12.10) into (5.12.4). The value of  $L_{r_{k+1}}$  is

$$\alpha^{r_k+1} (\alpha b_{r_{k+1}+1} - b_{r_{k+1}} - c) \\ + \frac{b_{r_{k+1}+1} - b_{r_{k+1}}}{b_{r_{k+1}+1}} \left\{ \sum_{r_{k+1} < i \leq r_k} \alpha^i (\alpha b_{i+1} - b_i - c) \right. \\ + \frac{b_{r_k}}{b_{r_{k-1}}} \sum_{r_{k-1} < i \leq r_{k-2}} \alpha^i (\alpha b_{i+1} - b_i - c) + \dots \quad (5.12.11) \\ \left. + \frac{b_{r_2} \cdot b_{r_4} \cdots b_{r_k}}{b_{r_1} \cdot b_{r_3} \cdots b_{r_{k-1}}} \sum_{r_1 < i \leq m-1} \alpha^i (\alpha b_{i+1} - b_i - c) \right\}.$$

To summarize: Relations for  $X_k$  are determined step by step, starting from  $X_{m-1}$ . For each  $X$ , we find that either (5.12.8) or (5.12.10) is satisfied. If (5.12.8) holds, the relation for the next  $X$  is determined after examining the sign of (5.12.9). If (5.12.10) holds, the relation for the next  $X$  is determined by examining the sign of (5.12.11).

The algorithm for the construction of the solution is now clear. By backward induction, a system of equations like (5.12.8) and (5.12.10) is satisfied by the  $X$ 's. After each new  $X$  is related to the preceding  $X$ 's, the previous equations are reduced, using the newly characterized  $X$ . The end result is a system of linear equations in the  $X$ 's with a triangular coefficient matrix, so that the explicit values of the  $X$ 's can be calculated recursively.

### 5.13 Problems.

1. (a) Two products  $A$  and  $B$  when manufactured must pass through four machine operations I, II, III, and IV. The machine time in hours per unit produced is as follows:

	I	II	III	IV
A	2	4	3	1
B	$\frac{1}{4}$	2	1	4

The total machine time available on machines I, II, III, and IV is 45, 100, 300, and 50 hours, respectively. Product  $A$  sells for \$6 per unit and product  $B$  for \$4 per unit. What combination of products  $A$  and  $B$  should be manufactured in order to maximize profit?

- (b) Suppose market demands are such that at least 22 units of product  $A$  are required. Solve the preceding problem with this constraint.

2. A gasoline refinery has two types of crude oil,  $A$  and  $B$ , which may be used to produce fuel oil and gasoline in any one of the following three patterns:

- (1) 1 unit of  $A$  + 2 units of  $B$   $\rightarrow$  2 units of fuel oil + 3 units of gasoline.
- (2) 2 units of  $A$  + 1 unit of  $B$   $\rightarrow$  5 units of fuel oil + 1 unit of gasoline.
- (3) 2 units of  $A$  + 2 units of  $B$   $\rightarrow$  2 units of fuel oil + 1 unit of gasoline.

Assuming that fuel oil sells for \$1 a unit, and gasoline sells for \$10 a unit, what is the most profitable production scheme if we start out with 10 units of  $A$  and 15 units of  $B$ ?

3. A man wishes to invest \$1000 in three common stocks, and no more than \$400 in any one of them. Stock  $A$  sells for \$50 a share and pays a dividend of \$2 a year; stock  $B$  sells for \$200 a share and pays a dividend of \$5 a year; stock  $C$  sells for \$20 a share and pays no dividend, but has an even chance (probability  $\frac{1}{2}$ ) of appreciating to \$25 a share in one year. If it does not appreciate to \$25, it will stay unchanged. What amount of

each stock should be bought to maximize dividends plus expected appreciation over one year? (Note: Fractional shares may be purchased.)

4. Solve the linear programming problem

$$\min (x_1 + 2x_2 + 3x_3 + \cdots + nx_n)$$

subject to the constraints

$$x_i \geq 0, \quad x_1 + x_2 + \cdots + x_i \geq i \quad (i = 1, \dots, n).$$

5. Construct an example in which there is no feasible vector satisfying (5.1.1).

6. Construct an example such that the constraint set (5.1.1) is nonvoid but such that Problem I has no solution.

7. There are  $n$  refineries with daily capacity  $D_i$  ( $i = 1, 2, \dots, n$ ) and  $m$  oilfields with daily production rate  $Q_j$  ( $j = 1, 2, \dots, m$ ). The value of sending one barrel of oil from the  $j$ th source to the  $i$ th refinery is  $p_{ij}$ . Assume that

$$\sum_{i=1}^n D_i \geq \sum_{j=1}^m Q_j.$$

Set up the linear programming problem determining the refining process on the daily oil yield which maximizes value subject to the natural constraints.

8. Prove that if  $\mathbf{b}$  is a vector with all components nonnegative and there exists an admissible vector  $\mathbf{y}$  for Problem II, then Problems I and II have solutions.

9. Prove that if  $\mathbf{b}$  is a vector with all components positive and  $\mathbf{A}$  is a matrix with all components positive, then Problems I and II have solutions.

10. (Generalized Weak Le Chatelier Principle.) Consider Problem I:  $\max_{\mathbf{x}} (\mathbf{c}, \mathbf{x})$  subject to the restrictions  $\mathbf{Ax} \leq \mathbf{b}$ ,  $\mathbf{x} \geq \mathbf{0}$ , and let  $\mathbf{x}_0$  denote a solution. Let  $\delta\mathbf{c}$  denote a variation of the vector  $\mathbf{c}$  so that  $(\mathbf{c} + \delta\mathbf{c}, \mathbf{x})$  is maximized under the same restrictions as before by  $\mathbf{x}_0 + \delta\mathbf{x}_0$ . Prove that  $\mathbf{x}_0$  satisfies  $(\delta\mathbf{c}, \delta\mathbf{x}_0) \geq 0$ .

11. An airplane is able to carry  $M$  pounds. Its pilot is asked to load  $n$  objects; the  $i$ th object weighs  $\alpha_i$  and its value is  $\beta_i$ . How should the plane be loaded so as to maximize the value of its cargo without exceeding the weight limit? Set up the problem mathematically and solve it.

12. Given two sets of numbers  $a_1, a_2, \dots, a_n$  and  $b_1, b_2, \dots, b_n$ , find the arrangement of  $a_i$  which minimizes

$$\sum_{i=1}^n b_i a_{\nu_i},$$

where  $\nu_i$  represents a permutation of the indices  $1, 2, \dots, n$ .

13. A firm must schedule the production of  $n$  objects, and the  $i$ th object takes  $T_i$  time units to process. The total waiting time of a single object is the length of time until its processing is completed. Find the order of processing the  $n$  objects which minimizes the total waiting time of all objects.

14. There are  $n$  objects to be processed. Let  $T_i$  represent the processing time of the  $i$ th object, and let  $D_i$  be the due date of the  $i$ th object. If the objects are processed in numerical order, the tardiness of the  $i$ th object is  $\delta_i = T_1 + T_2 + \dots + T_i - D_i$ . Let  $\Delta = \max \delta_i$ . Show that if  $D_1 \leq D_2 \leq D_3 \leq \dots \leq D_n$ , the order  $1, 2, 3, \dots, n$  minimizes  $\Delta$ .

15. Show that the system of linear equations  $\mathbf{yA} = \mathbf{a}$  has no nonnegative solution if and only if the inequalities  $\mathbf{Ax} \leq \mathbf{0}$  and  $(\mathbf{a}, \mathbf{x}) > 0$  have a solution.

16. Let  $\mathbf{A}$  and  $\mathbf{B}$  be two  $m \times n$  matrices, and  $b_{ij} \geq 0$ . We call a real number  $\lambda$  admissible if there exists  $\mathbf{x} > \mathbf{0}$  such that  $\mathbf{Ax} \geq \lambda \mathbf{Bx}$ . Let  $\Omega$  denote the set of all admissible  $\lambda$ , and assume that  $\Omega$  is nonvoid and bounded above.

We call  $\mu$  admissible if there exists  $\mathbf{y} > \mathbf{0}$  such that  $\mathbf{yA} \leq \mu \mathbf{yB}$ . Let  $M$  denote all admissible  $\mu$ , and assume that  $M$  is nonvoid and bounded below. Prove that

$$\lambda_0 = \sup_{\lambda \in \Omega} \lambda \geq \mu_0 = \inf_{\mu \in M} \mu.$$

17. Prove that if  $((\mathbf{B} + \mathbf{A})\mathbf{x}, \mathbf{y}) > 0$  for every  $\mathbf{x} > \mathbf{0}$  and  $\mathbf{y} > \mathbf{0}$ , then the assumptions of Problem 16 are satisfied and moreover  $\lambda_0 = \mu_0$ .

18. Let  $v(\mathbf{b}) = \max_{\mathbf{x}} (\mathbf{c}, \mathbf{x})$ , where  $\mathbf{x}$  satisfies  $\mathbf{x} \geq \mathbf{0}$ ,  $\mathbf{Ax} \leq \mathbf{b}$ . Prove that  $v(\mathbf{b})$  is a concave function of  $\mathbf{b}$ , and for the same function  $v(\mathbf{b})$  show that

$$\left( \frac{\partial v}{\partial b_k} \right)_+ \leq y_k \leq \left( \frac{\partial v}{\partial b_k} \right)_-$$

for any set of shadow prices  $y_1, y_2, \dots, y_m$  where  $+$  and  $-$  denote right- and left-hand derivatives respectively.

19. Let  $\mathbf{c}$  denote any vector of  $n$  components. Let  $I$  denote the indices  $i$  for which  $c_i > 0$ ,  $J$  the indices  $j$  for which  $c_j = 0$ , and  $K$  the indices  $k$  for which  $c_k < 0$ .

For  $i \in I$ , define  $\alpha_i = \langle 0, 0, \dots, 1, 0, \dots, 0 \rangle$ , with 1 only in the  $i$ th component; for  $j \in J$ , define  $\beta_j = \langle 0, 0, \dots, 1, 0, \dots, 0 \rangle$ , with 1 only in the  $j$ th component; and for  $i \in I, k \in K$ , define

$$\nu_{ik} = \left\langle 0, \dots, \frac{1}{c_i}, 0, \dots, -\frac{1}{c_k}, 0, \dots, 0 \right\rangle,$$

with  $1/c_i$  in the  $i$ th component,  $-1/c_k$  in the  $k$ th component, and zero elsewhere.

Let  $T(c)$  denote the matrix whose column vectors are  $\alpha$ 's,  $\beta$ 's, and  $\nu$ 's. Prove that  $(c, x) \geq 0, x \geq 0$  if and only if  $x = T(c)\omega$  for some  $\omega \geq 0$ .

20. Let  $x_0$  and  $y_0$  represent solutions to Problems I and II, respectively, and let  $x_0 + \delta x_0$  and  $y_0 + \delta y_0$  represent solutions to Problems I and II where  $c$ ,  $b$ , and  $A$  are replaced by  $c + \delta c$ ,  $b + \delta b$ , and  $A + \delta A$ , respectively. Prove that

$$-(\delta c, \delta x_0) + (y_0 + \delta y_0, \delta A \delta x_0) + (\delta y_0, A \delta x_0) \leq 0$$

and

$$-(\delta y_0, \delta b) + (\delta y_0, \delta A(x_0 + \delta x_0)) + (\delta y_0, A \delta x_0) \geq 0.$$

(Hint: Use the saddle-value theorem, Theorem 5.3.1.)

21. Let  $S$  be a bounded, closed, polyhedral, convex set in  $E^n$  such that  $y > 0$  for any  $y$  in  $S$ . Prove that there exists a vector  $v$  such that  $v \gg 0$  and  $(v, y) \leq 0$  for all  $y$  in  $S$ .

Problems 22–26 are based on the following structure. Let  $A_1, A_2, \dots, A_m$  be given vectors in  $n$ -space, with coordinates consisting of 1's and 0's such that (a) the 1's in each  $A_i$  are consecutive and each  $A_i$  contains at least one 1; (b) if the coordinates in  $A_i$  which are 1 are  $a, a+1, \dots, b$  and the coordinates in  $A_j$  which are 1 are  $c, c+1, \dots, d$ , and if  $a < c$ , then  $b \leq d$ .

Let  $r_i$  ( $i = 1, \dots, n$ ) denote positive constants and  $t_i$  ( $i = 1, \dots, n$ ) given constants. Consider Problem A: Minimize

$$\lambda X + \sum_{i=1}^n z_i$$

with respect to  $X$  and  $z_i$ , subject to the constraints

$$(A_i, z) \geq t_i - X \quad (i = 1, \dots, n),$$

$$0 \leq z_i \leq r_i \quad (i = 1, \dots, n),$$

$$\lambda \geq 0.$$

22. Show that the caterer's problem of Section 5.11 is a special case of Problem A.

Assume in what follows that the constraint set is nonvoid.

23. For each fixed  $X$  describe a method for solving Problem A.

24. If  $z(X) = z_1^0(X) + \dots + z_n^0(X)$  is the solution of Problem A for fixed  $X$ , show that  $z(X)$  is a convex, nonincreasing, polygonal function of  $X$ .

25. Solve Problem A completely if  $A_i = \langle 0, 0, \dots, 1, 0, \dots, 0 \rangle$ , with 1 occurring in the  $i$ th coordinate only.

26. On the basis of Problems 23 and 24, describe an algorithm for the solution of the general Problem A.

## NOTES AND REFERENCES TO CHAPTER 5

§ 5.1. The formal theory of linear programming originated shortly after World War II, although special linear programming models have been considered over the past century. For example, results in the theory of linear inequalities can be traced back to the early works of Weyl [253], Farkas [79], and Motzkin [186]. See [163] for complete references.

Dantzig is commonly called the father of linear programming because it was he, in conjunction with von Neumann, who discovered and popularized the basic computational algorithm for solving linear programming problems. More recently, Charnes and Cooper have made a vast number of important contributions to the theory and its applications. An introduction to the theory and the simplex algorithm of linear programming is given in Charnes, Cooper, and Henderson [46].

An intermediate-level linear programming textbook by Gass [103] contains an account of many computing devices useful for different types of linear programming problems, and an extensive bibliography of linear programming applications.

Students with a moderate undergraduate mathematical training but oriented toward economic theory are referred to Dorfman, Samuelson, and Solow [65].

Among the numerous other practical linear programming texts are Ferguson and Sargent [82] and Reinfeld and Vogel [203].

The development of linear programming theory has also been substantially influenced by the study of activity analysis models, notably by Koopmans and the economists of the Cowles Commission. The reader is referred to two volumes on activity analysis edited by Koopmans [146], and also to Morgenstern [182] and Dorfman [64].

The first linear programming problem solved, although it was not so labeled, was a version of the transportation problem (Hitchcock [124]). The method used was a form of the simplex method. The earliest modern linear programming problem considered was the diet problem, which was first formulated and analyzed by Stigler [229]. Also of historical interest are the programming models of air force operations as set forth by Wood and Dantzig.

The most elaborate large-scale linear programming studies have been concerned with the workings of the oil industry. A review of the difficulties and accomplishments of linear programming in this industry is given in Manne [171].

The most common linear programming problem is the transportation problem, sometimes referred to as the Hitchcock distribution problem. There is an extensive literature devoted to the transportation problem and its variants (see, for example, Kantorovitch [131], Koopmans and Reiter [150], Flood [84], and Ford and Fulkerson [87]). The warehouse storage model cited in this chapter is due to Charnes and Cooper [45]. Various forms of activity analysis problems have been dealt with by Leontief [166], Samuelson [210], and Koopmans [146].

The methods and theory of linear programming have become a basic tool for operations researchers. Many operations research texts incorporate introductory summaries of the methodology of linear programming (see Chapters 11-13 of Churchman, Ackoff, and Arnoff [50] and Vajda [236]).

**§ 5.2.** The interpretation of the dual linear programming problem is not too clearly understood. For variations of the meanings ascribed here to the dual solution, see Samuelson [212] and Uzawa [234] (see also Koopmans [147]).

**§ 5.3.** The proof of Lemma 5.3.1 is new and represents a slight extension of the usual basis for deriving the duality and existence theorems of linear programming.

Theorem 5.3.1 is a special case of what is now known as the classical Kuhn-Tucker theorem of concave programming [164].

**§ 5.4.** The duality and existence theorems are the principal theorems of linear programming; almost all other developments rest on these two propositions. Numerous slightly different versions of the two theorems have been published, notably by Dantzig [55] and Gale, Kuhn, and Tucker [99].

**§ 5.5.** The equivalence of matrix-game problems and linear programming problems was recognized first by Dantzig [55] and von Neumann. The interaction between these two mathematical disciplines has accelerated the growth of each. What is more important is the common mode of thinking and methodology which underlies both subjects.

**§ 5.6.** The matrix extensions of the duality and existence theorems of linear programming were first suggested by Gale, Kuhn, and Tucker [99]. They proved Theorem 5.6.2 as well as other generalizations. The proof of that theorem given here is new, and the proof of Theorem 5.6.1, as well as its formulation, appears also to be new.

**§ 5.7.** The formulation and method of solution of the warehouse storage problem are due to Charnes and Cooper [45]. They did not give a rigorous proof of the optimality of the algorithm proposed for solving the problem. The formal arguments are indicated here. Charnes and Cooper have pursued many extensions of these models.

**§ 5.8.** The optimal assignment problem with all its ramifications was investigated by von Neumann [240]. The more elegant treatment of these topics followed in this section was suggested by Dantzig.

**§ 5.9.** More than half of the applications of linear programming described in the proceedings of a 1956 conference on the subject [167] were to transportation problems. Journals interested in programming have printed innumerable refinements and extensions of versions of the model described in this section. See Ford and Fulkerson [85, 86] and Kuhn [158] as typical.

**§ 5.10.** The maximal-flow, minimal-cut theorem of network flow is a beautiful application of the duality theorem. This result was discovered and proved by Ford and Fulkerson [85]. At about the same time, other writers showed how to treat many set-theoretic combinatorial problems of mathematics as applications of the duality theorem. For example, let  $b$  be a finite partially ordered set consisting of elements  $b_1, \dots, b_m$ . A chain of  $P$  is a set of one or more distinct  $b_{i_1}, \dots, b_{i_k}$  with  $b_{i_1} \geq b_{i_2} \geq \dots \geq b_{i_k}$ . A set of disjunct chains covers  $P$  if every element of  $P$  belongs to some chain of the set. Find a minimal covering of  $P$ . (Dantzig and Hoffman [58].) For other applications of the duality theorem, see [90] and [87].

**§ 5.11.** The caterer's problem was proposed and partially solved by Jacobs [129]. A different attack on the general problem, yielding an algorithm for the solution, was undertaken by Gaddum, Hoffman, and Sokolowsky [92]. Their methods apply to a more general type of problem (see Problems 22–26).

Prager [199] recognized that the caterer's problem could be formulated as a transportation problem and was therefore susceptible to the numerous specialized procedures developed for the transportation problem.

**§ 5.12.** The price speculation model is due to Arrow and Karlin [15].

**§ 5.13.** The result of Problem 10 was first noted by Samuelson. An extension of this result, embodied in the statement of Problem 20, is due to Beckmann. The assertions of Problems 13 and 14 are taken from Smith [226]. Problem 18 is due to Samuelson [212]. Problem 19 was suggested by Uzawa (unpublished). Problems 22–26 are based on Gaddum, Hoffman, and Sokolowsky [92].

## CHAPTER 6

### COMPUTATIONAL METHODS FOR LINEAR PROGRAMMING AND GAME THEORY

We have seen that a wide variety of problems can be cast in game form if not in linear programming form. To one grappling with a specific problem, however, there is nothing so important as a numerical answer. In this chapter we discuss relatively simple, efficient procedures for obtaining numerical answers to linear programming and game problems. We concentrate on two such procedures: the Dantzig simplex method for solving linear programming problems and the Brown algorithm for determining the value of the game.

Both of these procedures have been programmed for machine computation, and are said to work far more efficiently than might be expected in the light of their known theoretical bounds. In many respects the practical importance of linear programming and game theory depends on these numerical procedures.

The first two sections of this chapter are devoted to a description and analysis of the simplex method; the third section illustrates the workings of this method in a product mix problem. As a sample of some other procedures used in connection with special linear programming problems, we present in Section 6.4 a technique for solving network flow problems. The final sections discuss the Brown algorithm and a differential-equations method of approximating the value of a game.

This chapter is intended only as an introduction to the computational methods used in solving decision problems. It should be remembered that often the special structure of a problem, in conjunction with the general philosophy of decision theory, leads more directly to its solution than the routine application of the standard computational procedure. This is necessarily true of nonlinear programming problems (see Chapter 7), for which no general numerical methods are available. Consequently, this chapter emphasizes the interaction between general theory and computational procedure.

**6.1 The simplex method.** Consider the linear programming problem I: Find  $\mathbf{x}$  which maximizes  $L(\mathbf{x}) = (\mathbf{c}, \mathbf{x})$ , with  $\mathbf{x}$  subject to the constraints

$$\sum_{j=1}^n a_{ij}x_j \leq b_i \quad (i = 1, \dots, m), \quad x_j \geq 0. \quad (6.1.1)$$

The set of  $\mathbf{x}$  satisfying the constraints will be denoted by  $S$ .

We now describe a computational procedure—the simplex method—for obtaining a solution  $\mathbf{x}^0$  to Problem I. (The simplex algorithm for the dual problem is presented in Section 6.2.) Justifications of the various steps of the procedure are given in the course of the description.

The method consists, essentially, of a systematic way of determining the relevant extreme points of the constraint set  $S$ . Since the function being maximized is linear and the constraint set  $S$  is convex, it is clear that the maximum, if it exists, must be achieved for at least one extreme point of  $S$ . The simplex method begins by first determining any extreme point. In the case in which the constraints are described by linear inequalities, this is easy to do (see paragraph E of this section). Once an extreme point of  $S$  is known, a test can be applied to determine whether this is a solution. If it is not, we seek a new extreme point of  $S$ , one that is more likely to maximize the given linear form  $L$ . The direction in  $S$  in which we move to locate this new extreme point is in essence the direction of the gradient of  $L$ . After a finite number of steps either a solution is obtained, or we learn that the maximum is infinite and the problem has no solution.

The detailed elaboration of the procedure follows.

*A. Introduction of “slack” variables.* The first step of the simplex method is to convert the inequalities of (6.1.1) into equalities. This is done by introducing  $m$  new variables,  $x_{n+1}, \dots, x_{n+m}$ , which are known as “slack” variables and are defined by

$$x_{n+k} = b_k - \sum_{j=1}^n a_{kj}x_j \quad (k = 1, \dots, m). \quad (6.1.2)$$

If we convert the matrix  $\mathbf{A}$  into the matrix

$$\tilde{\mathbf{A}} = \left\| \begin{array}{cccccc} a_{11} & \dots & a_{1n} & \overbrace{1 \dots 0}^m \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & \dots & a_{mn} & 0 \dots 1 \end{array} \right\|,$$

and  $\mathbf{x}$  into  $\tilde{\mathbf{x}} = \langle x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m} \rangle$ , the system of constraints  $\mathbf{Ax} \leq \mathbf{b}$ ,  $\mathbf{x} \geq \mathbf{0}$ , in conjunction with (6.1.2), can be written as

$$\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{b}, \quad \tilde{\mathbf{x}} \geq \mathbf{0}. \quad (6.1.3)$$

If we now replace  $\mathbf{c}$  by  $\tilde{\mathbf{c}} = \langle c_1, c_2, \dots, c_n, c_{n+1}, \dots, c_{n+m} \rangle$ , where  $c_{n+1} = \dots = c_{n+m} = 0$ , the linear programming problem becomes:

Find  $\mathbf{x}_0$  such that

$$(\mathbf{c}, \mathbf{x}_0) = \max_{\tilde{\mathbf{x}}} (\mathbf{c}, \mathbf{x})$$

among all  $\mathbf{x}$  vectors satisfying the conditions  $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{b}$  and  $\mathbf{x} \geq \mathbf{0}$ .

The computational procedure deals only with this equivalent problem. Henceforth, we drop the tildes in (6.1.3) and assume that our original constraints for  $\mathbf{A}$  were as described in (6.1.3). Restating the problem in new notation, we have: Find a solution to  $\max_{\lambda} (\mathbf{c}, \lambda)$ , subject to the conditions  $\mathbf{A}\lambda = \mathbf{b}$  and  $\lambda \geq \mathbf{0}$ , where  $\mathbf{A}$  is an  $m \times n$  matrix.

Let  $\mathbf{p}_j$  denote the  $j$ th column vector of  $\mathbf{A}$ , and let  $\mathbf{p}_0$  denote the vector  $\mathbf{b}$ . Each  $\mathbf{p}_j$  has  $m$  components.

In order to avoid certain inessential technical difficulties associated with the simplex method, we tentatively make the following assumption.

**ASSUMPTION 6.1.1.** The vector  $\mathbf{p}_0$  cannot be expressed as a positive combination of fewer than  $r$  of the  $\mathbf{p}_j$ , where  $r$  is the rank of  $\mathbf{A}$ .

In most cases the rank of  $\mathbf{A}$  is equal to its number of rows. This is always so when the original constraint set is composed entirely of inequalities and  $\mathbf{A}$  is the result of enlarging the prescribed constraint matrix to include slack variables. In this circumstance the submatrix of  $\mathbf{A}$  whose columns correspond to the slack variables is nonsingular. From here on, in order to simplify the exposition, we assume that  $\mathbf{A}$  is of rank  $m$ , the number of its rows. We leave to the reader the task of modifying the procedures for the case in which  $\mathbf{A}$  is of rank smaller than  $m$ .

Assumption 6.1.1 is known as the nondegeneracy assumption. After presenting a method for finding a solution under this assumption, we shall show how to circumvent it and still obtain a solution.

(When degeneracies are present, a lack of uniqueness appears in the process of obtaining new extreme points and a certain degree of arbitrariness is introduced. This is primarily a nuisance rather than a fundamental difficulty: it means that the procedure must be carried out with greater care and that some of its simplicity is sacrificed.)

*B. Characterization of extreme points.* Let  $\Lambda$  denote the set of admissible  $\lambda$ 's which satisfy the constraints

$$\lambda_1 \mathbf{p}_1 + \lambda_2 \mathbf{p}_2 + \cdots + \lambda_n \mathbf{p}_n = \mathbf{p}_0, \quad \lambda_i \geq 0. \quad (6.1.4)$$

► **LEMMA 6.1.1.** If  $\lambda$  is an extreme point of  $\Lambda$ , then  $\lambda$  has at most  $m$  non-zero components.

*Proof.* Suppose to the contrary that  $\lambda_1, \lambda_2, \dots, \lambda_{m+1} > 0$ . Consider the matrix  $\mathbf{A} = \|\mathbf{p}_1, \dots, \mathbf{p}_{m+1}\|$  and the system of homogeneous linear

equations  $\mathbf{A}\mathbf{z} = \mathbf{0}$ . Since the number of unknowns clearly exceeds the rank of  $\mathbf{A}$ , there exists a nontrivial solution  $\mathbf{z}^0 = \langle z_1^0, \dots, z_{m+1}^0 \rangle$ . Enlarge  $\mathbf{z}^0$  to  $\tilde{\mathbf{z}}^0 = \langle z_1^0, \dots, z_{m+1}^0, 0, \dots, 0 \rangle$  by adding  $n - m - 1$  zeros. Select  $\epsilon > 0$  sufficiently small so that the components of the vectors

$$\lambda^1 = \lambda + \epsilon \tilde{\mathbf{z}}^0 \quad \text{and} \quad \lambda^2 = \lambda - \epsilon \tilde{\mathbf{z}}^0$$

are all nonnegative. Since  $\mathbf{A}\lambda^1 = \mathbf{b}$  and  $\mathbf{A}\lambda^2 = \mathbf{b}$ , it follows that  $\lambda^1, \lambda^2 \in \Lambda$ . But  $\lambda^1 \neq \lambda^2$ , and

$$\lambda = \frac{1}{2}(\lambda^1 + \lambda^2).$$

This contradicts the assumption that  $\lambda$  is an extreme point, and the lemma follows.

Assumption 6.1.1 was not used in this lemma, but is used in the one that follows.

►LEMMA 6.1.2. If  $\lambda \in \Lambda$  and  $\lambda_1, \dots, \lambda_m > 0, \lambda_{m+1} = \dots = \lambda_n = 0$ , then  $\lambda$  is an extreme point of  $\Lambda$ .

*Proof.* Suppose that there exist  $\lambda^1, \lambda^2 \in \Lambda$  such that

$$\lambda = \frac{1}{2}(\lambda^1 + \lambda^2).$$

Since the last  $n - m$  components of  $\lambda$  are equal to zero, the last  $n - m$  components of both  $\lambda^1$  and  $\lambda^2$  must equal zero. Then  $\mathbf{A}\tilde{\lambda}^1 = \mathbf{p}_0$  and  $\mathbf{A}\tilde{\lambda}^2 = \mathbf{p}_0$ , where  $\tilde{\lambda}^i = \langle \lambda_1^i, \lambda_2^i, \dots, \lambda_m^i \rangle$ . Thus  $\mathbf{A}(\tilde{\lambda}^1 - \tilde{\lambda}^2) = \mathbf{0}$ . If  $\tilde{\lambda}^1 - \tilde{\lambda}^2 \neq \mathbf{0}$ , then  $\epsilon$  can be chosen sufficiently small so that

$$\mu = \tilde{\lambda}^1 + \epsilon(\tilde{\lambda}^1 - \tilde{\lambda}^2) \geq \mathbf{0} \quad \text{and} \quad \mu_1 \mathbf{p}_1 + \mu_2 \mathbf{p}_2 + \dots + \mu_n \mathbf{p}_n = \mathbf{p}_0.$$

If some component of  $\tilde{\lambda}^1 - \tilde{\lambda}^2$  is negative, we can increase  $\epsilon$  until at least one of the components of  $\mu$  vanishes while  $\mu$  persists as a nonnegative vector. We thus obtain a representation of  $\mathbf{p}_0$  by fewer than  $m$  of the  $\mathbf{p}_i$ , contradicting Assumption 6.1.1. If some component of  $\tilde{\lambda}^1 - \tilde{\lambda}^2$  is positive, we choose  $\epsilon$  negative and a similar argument applies. Hence in all cases  $\lambda^1 = \lambda^2$ , and thus  $\lambda$  is an extreme point.

C. *Getting a new extreme point from the old.* Suppose that  $\lambda^0$  has  $\lambda_1^0, \dots, \lambda_m^0 > 0$  and  $\lambda_{m+1}^0 = \dots = \lambda_n^0 = 0$  and satisfies (6.1.4). (We shall discuss later how to obtain such a point.) Then  $\lambda^0$  is an extreme point of  $\Lambda$  (Lemma 6.1.2). It follows that the vectors  $\mathbf{p}_1, \dots, \mathbf{p}_m$  are linearly independent and hence constitute a basis of  $m$ -space. Otherwise, there exists a nonzero vector  $\mu = \langle \mu_1, \dots, \mu_m \rangle$  such that

$$\sum_{i=1}^m \mu_i \mathbf{p}_i = \mathbf{0}.$$

Then

$$\lambda^0 = \frac{(\lambda^0 + \epsilon\mu) + (\lambda^0 - \epsilon\mu)}{2}$$

and  $\lambda^0 \pm \epsilon\mu$  are nonzero distinct feasible vectors for  $\epsilon$  sufficiently small, contradicting the extremal property of  $\lambda^0$ . Consequently, for every vector  $\mathbf{p}_i$  ( $i = 1, 2, \dots, n$ ) there exist numbers  $x_{ji}$  ( $j = 1, 2, \dots, m$ ) such that

$$\mathbf{p}_i = \sum_{j=1}^m \mathbf{p}_j x_{ji}.$$

We refer to  $\mathbf{X} = \|x_{ji}\|$  ( $j = 1, \dots, m$ ;  $i = 1, \dots, n$ ) as the “tableau” associated with the extreme-point vector  $\lambda^0$ . The corresponding basis vectors  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m$  used in the representation of  $\mathbf{p}_0$  with weights  $\lambda_j^0$  are called the basis vectors of the extreme point  $\lambda^0$ .

We may obviously rewrite (6.1.4) as

$$\lambda_1^0 \mathbf{p}_1 + \dots + \lambda_m^0 \mathbf{p}_m + \theta \left( \mathbf{p}_k - \sum_{j=1}^m \mathbf{p}_j x_{jk} \right) = \mathbf{p}_0,$$

where  $k$  denotes an arbitrary index between  $m + 1$  and  $n$ . This equation may be rearranged to read

$$\sum_{j=1}^m (\lambda_j^0 - \theta x_{jk}) \mathbf{p}_j + \theta \mathbf{p}_k = \mathbf{p}_0. \quad (6.1.5)$$

So long as the coefficients of  $\mathbf{p}_k$  and  $\mathbf{p}_j$  ( $j = 1, \dots, m$ ) in (6.1.5) are non-negative and  $\theta$  is steadily increased from zero, we move in a direction of  $\Lambda$  where  $\mathbf{p}_k$  is increasingly involved in the representation of  $\mathbf{p}_0$ . Let  $\lambda^1$  denote the point of  $\Lambda$  that yields the representation of  $\mathbf{p}_0$  given in (6.1.5).

Consider for  $k$  fixed

$$\begin{aligned} (\mathbf{c}, \lambda^1) &= \sum_{j=1}^m (\lambda_j^0 - \theta x_{jk}) c_j + \theta c_k \\ &= (\mathbf{c}, \lambda^0) + \theta \left( c_k - \sum_{j=1}^m x_{jk} c_j \right) \\ &= (\mathbf{c}, \lambda^0) - \theta (z_k - c_k), \end{aligned} \quad (6.1.6)$$

where  $z_k$  is implicitly defined by the last equation. We now distinguish three cases.

*Case 1.* There exists a value  $k_0$  ( $m + 1 \leq k_0 \leq n$ ) for which  $z_{k_0} - c_{k_0} < 0$ , where

$$z_k = \sum_{j=1}^m x_{jk} c_j$$

and  $x_{jk_0} \leq 0$  for  $j = 1, \dots, m$ . In this case  $\lambda_j^0 - \theta x_{jk_0} \geq 0$ , and hence  $\lambda^1$  qualifies as a member of  $\Lambda$  for every  $\theta > 0$ . Letting  $\theta$  (the value of  $\lambda_{k_0}$ ) tend to  $+\infty$ , we see that  $(\mathbf{c}, \lambda^1)$  increases beyond all limits and the problem has no solution; i.e., the maximum is infinite.

*Case 2.* There exists a  $k_0$  ( $m + 1 \leq k \leq n$ ) for which  $z_{k_0} - c_{k_0} < 0$  and  $x_{jk_0} > 0$  for some  $j$ . Let

$$\theta = \min_j \frac{\lambda_j^0}{x_{jk_0}},$$

the minimum being taken over all  $j$  for which  $x_{jk_0} > 0$ . Clearly,  $\theta > 0$  and  $\lambda_j^0 - \theta x_{jk_0} \geq 0$  for  $j = 1, \dots, m$ , with equality valid for at least one  $j$ . Two such terms cannot vanish together, however; if they do,  $\mathbf{p}_0$  is expressed by a positive combination of fewer than  $m$  of the  $\mathbf{p}_j$ , in contradiction to Assumption 6.1.1. When two components simultaneously vanish, we say that a degeneracy occurs. With  $\theta$  determined as indicated, the components of  $\lambda^1$  are  $\lambda_j = \lambda_j^0 - \theta x_{jk_0}$  for  $j = 1, \dots, m$ ,  $\lambda_{k_0} = \theta$ , and zeros elsewhere. The vector  $\lambda^1$  possesses  $m$  nonzero components and therefore, by Lemma 6.1.2, is an extreme point of  $\Lambda$ . Moreover,  $(\mathbf{c}, \lambda^1) > (\mathbf{c}, \lambda^0)$ , so that the extreme point  $\lambda^1$  definitely yields a larger value than  $\lambda^0$ .

*Case 3.* If  $z_k - c_k \geq 0$  for all  $k$ , then  $\lambda^0$  is optimal. To prove this statement, we must show that

$$\sum_{i=1}^n \lambda_i \mathbf{p}_i = \mathbf{p}_0$$

implies  $(\lambda, \mathbf{c}) \leq (\lambda^0, \mathbf{c})$ . Note that  $z_k = c_k$  for  $1 \leq k \leq m$ . From the condition of Case 3, it follows that

$$\sum_{i=1}^n \lambda_i c_i \leq \sum_{i=1}^n \lambda_i \sum_{j=1}^m x_{ji} c_j = \sum_{j=1}^m c_j \sum_{i=1}^n \lambda_i x_{ji}.$$

The proof will be finished if it is shown that

$$\lambda_j^0 = \sum_{i=1}^n \lambda_i x_{ji}.$$

Explicitly,

$$\mathbf{p}_0 = \sum_{i=1}^n \lambda_i \mathbf{p}_i = \sum_{i=1}^n \lambda_i \sum_{j=1}^m x_{ji} \mathbf{p}_j = \sum_{i=1}^m \mathbf{p}_j \sum_{i=1}^n \lambda_i x_{ji} = \sum_{j=1}^m \lambda_j^0 \mathbf{p}_j.$$

By virtue of Assumption 6.1.1 we deduce the equality of the coefficients, and the proof is complete.

*D. Formalization.* Suppose that we have an extreme point  $\lambda^0$  and a basis  $\mathbf{p}_1, \dots, \mathbf{p}_m$ . Compute  $x_{ik}$  of the following matrix, where

$$\mathbf{p}_k = \sum_{j=1}^m \mathbf{p}_j x_{jk} \quad (k = 0, 1, \dots, n),$$

and the auxiliary quantities are as indicated:

	$p_1$	$p_2$		$p_k$		$p_n$	$p_0$
$p_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1k}$		$\dots$	$x_{1n}$
$p_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2k}$		$\dots$	$x_{2n}$
$\vdots$							
$p_m$	$x_{m1}$	$x_{m2}$	$\dots$	$x_{mk}$		$\dots$	$x_{mn}$
$z:$	$z_1$	$z_2$	$\dots$	$z_k = \sum_{j=1}^m x_{jk} c_j$		$\dots$	$z_n$
$z - c:$	$z_1 - c_1$	$z_2 - c_2$	$\dots$	$z_k - c_k$		$\dots$	$z_n - c_n$

*Case 1.* If  $z_i - c_i \geq 0$  for all  $i$ , then  $\lambda^0 = \langle \lambda_1^0, \lambda_2^0, \dots, \lambda_m^0 \rangle$  is the solution of the maximization problem.

*Case 2.* If  $z_{k_0} - c_{k_0} < 0$  for some  $k_0$ , and  $x_{ik_0} \leq 0$  for all  $i$ , then  $\max(\mathbf{c}, \lambda) = \infty$ , the maximum being extended over all feasible  $\lambda$ .

*Case 3.* If  $z_{k_0} - c_{k_0} < 0$  for some  $k_0$ , and  $x_{ik_0} > 0$  for some  $i$ , then we can get a new extreme point  $\lambda^1$  for which  $(\mathbf{c}, \lambda^1) > (\mathbf{c}, \lambda^0)$ . In practice it is best to choose the index  $k_0$  for which  $z_{k_0} - c_{k_0}$  has the largest possible negative value; this tends to decrease the number of iterations needed for the simplex technique.

The components of  $\lambda^1$  are  $\lambda_i^1 = \lambda_i^0 - \theta x_{ik_0}$  for  $i = 1, \dots, m$ ,

$$\lambda_{k_0} = \theta = \min_{\substack{i \\ x_{ik_0} > 0}} \frac{\lambda_i^0}{x_{ik_0}},$$

and zeros elsewhere. Corresponding to this extreme point are a new basis and a new tableau.

Define  $l_0$  such that  $\lambda_{l_0}^0 - \theta x_{l_0, k_0} = 0$ . Then the basis vectors associated with  $\lambda^1$  are  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{l_0-1}, \mathbf{p}_{l_0+1}, \dots, \mathbf{p}_m, \mathbf{p}_{k_0}$ . In other words,  $\mathbf{p}_{k_0}$  has replaced  $\mathbf{p}_{l_0}$ .

The tableau  $\mathbf{Y} = \|y_{ji}\|$  of  $\lambda^1$  may be readily computed. To this end, suppose for simplicity that  $l_0 = 1$ . Then the new basis consists of the vectors  $\mathbf{p}_2, \dots, \mathbf{p}_m, \mathbf{p}_{k_0}$ .

Now

$$\mathbf{p}_i = \sum_{j=2}^m y_{ji} \mathbf{p}_j + y_{k_0 i} \mathbf{p}_{k_0}. \quad (6.1.7)$$

These  $y_{ji}$  can be expressed in terms of old  $x_{ji}$ . Indeed, since  $x_{1k_0}$  cannot be zero by Assumption 6.1.1, we obtain

$$\begin{aligned} \mathbf{p}_1 &= - \sum_{j=2}^m \frac{x_{jk_0}}{x_{1k_0}} \mathbf{p}_j + \frac{1}{x_{1k_0}} \mathbf{p}_{k_0}, \\ \mathbf{p}_i &= -x_{1i} \sum_{j=2}^m \frac{x_{jk_0}}{x_{1k_0}} \mathbf{p}_j + \frac{x_{1i}}{x_{1k_0}} \mathbf{p}_{k_0} + \sum_{j=2}^m x_{ji} \mathbf{p}_j \quad (i \geq 2). \end{aligned}$$

On comparison with (6.1.7), it follows that

$$y_{ji} = \begin{cases} x_{ji} - \frac{x_{1i}x_{jk_0}}{x_{1k_0}} & (j \neq k_0), \\ \frac{x_{1i}}{x_{1k_0}} & (j = k_0). \end{cases} \quad (6.1.8)$$

For the general case in which  $\mathbf{p}_{k_0}$  replaces  $\mathbf{p}_{r_0}$ , the index 1 in (6.1.8) is to be replaced throughout by  $r_0$ . Once a tableau  $\mathbf{Y}$  is explicitly known for the new extreme point  $\lambda^1$ , we are ready to repeat the process and get a further extreme point, one which yields a still larger value of  $(\mathbf{c}, \lambda)$ , unless we find that the problem has no solution or that  $\lambda^1$  is the solution. After a finite number of steps, since there are only a finite number of distinct extreme points, we obtain an optimal extreme vector.

*E. Obtaining an initial extreme point.* For the linear programming problem I we have the restrictions  $\mathbf{A}\lambda \leq \mathbf{b}$ ,  $\lambda \geq \mathbf{0}$ . The expanded matrix  $\tilde{\mathbf{A}}$  including the slack variables is

$$\left\| \begin{array}{cc|cc} a_{11} & \dots & a_{1n} & \overbrace{1 \dots 0}^m \\ \cdot & & \cdot & \cdot 1 \cdot \\ \cdot & & \cdot & \cdot \dots \\ \cdot & & \cdot & \cdot \dots \\ a_{m1} & \dots & a_{mn} & 0 \dots 1 \end{array} \right\| = \|\mathbf{A}, \mathbf{I}\|.$$

When  $\mathbf{b}$  is a nonnegative vector, the last  $m$  columns of  $\tilde{\mathbf{A}}$  constitute a basis of the extreme point  $\langle 0, \dots, 0, b_1, \dots, b_m \rangle$ .

This is not the case if some of the components of  $\mathbf{b}$  are negative, since then the linear combination of the vectors  $\mathbf{p}_{n+1}, \mathbf{p}_{n+2}, \dots, \mathbf{p}_{n+m}$  which

represent  $\mathbf{p}_0$  has coefficients  $\langle b_1, b_2, \dots, b_m \rangle$  not all nonnegative. In this case, to obtain an initial extreme point we introduce the concept of "artificial variables." Suppose for definiteness that the first  $r$  components  $b_1, b_2, \dots, b_r$  are negative and the remaining  $b$ 's are nonnegative. Consider the augmented matrix  $\mathbf{A}^* = \|\mathbf{A}, \mathbf{I}_m, \mathbf{P}_r\|$  with

$$\mathbf{P}_r = \left\| \begin{array}{ccc|c} -1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \\ \vdots & \ddots & \vdots & \\ 0 & -1 & & \\ 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & \end{array} \right\|$$

that is,  $\mathbf{P}_r$  is an  $m \times r$  matrix such that  $p_{ii} = -1$  ( $1 \leq i \leq r$ ) and  $p_{ij} = 0$  otherwise. We choose  $c_{n+m+1} = c_{n+m+2} = \dots = c_{n+m+r} = -M$ , with  $M$  positive and sufficiently large so that the last  $r$  components of  $\lambda$  in the maximization of  $(\lambda, \mathbf{c})$  are equal to zero, provided a solution exists.

For the enlarged matrix  $\mathbf{A}^*$  the last  $m$  columns constitute a basis and thus we secure an initial extreme point. It is standard practice in implementing the computational algorithm to keep the vector  $\mathbf{b}$  nonnegative. This means simply that suitable rows of the constraining equations are to be multiplied by  $-1$ , an operation that does not alter the choice of the starting basis.

This device of adding artificial variables is always applicable if no other initial extreme point is available. Its disadvantage is that it increases the number of iterations required by the simplex technique.

*F. Circumventing the nondegeneracy assumption.* It may be possible to represent  $\mathbf{p}_0$  by fewer than  $m$  vectors  $\mathbf{p}_i$ . In this case

$$\theta = \min_j \frac{\lambda_j^0}{x_{jk_0}}$$

may be achieved for two or more indices. It is conceivable at this point that a cycling among several extreme points occurs as the change of basis is effected. More specifically, if  $\theta$ 's of equal value are found, then on subsequent iterations one cannot be sure that the new extreme vector for a new basis furnishes a larger maximum.

This difficulty may be overcome by modifying the initial problem as

follows: Keep the vectors  $\mathbf{p}_i$  unchanged but replace  $\mathbf{p}_0$  by

$$\mathbf{p}_0(\epsilon) = \mathbf{p}_0 + \sum_{j=1}^n \epsilon^j \mathbf{p}_j,$$

where  $\epsilon$  is an unspecified positive parameter and sufficiently small. The method to be described is independent of the parameter  $\epsilon$  and in essence involves only slight additional care in applying the basic simplex algorithm. Let us consider a basis with index set  $I$  and tableau matrix  $\mathbf{X}$  (obviously all basis systems for the original and modified problems coincide). If

$$\sum_{i \in I} \lambda_i \mathbf{p}_i = \mathbf{p}_0,$$

then

$$\sum_{i \in I} \left( \lambda_i + \sum_{j=1}^n \epsilon^j x_{ij} \right) \mathbf{p}_i = \mathbf{p}_0(\epsilon).$$

Hence

$$\lambda'_r(\epsilon) = \lambda_r + \epsilon^r + \sum_{j \notin I} \epsilon^j x_{rj}.$$

Each  $\lambda'_r(\epsilon)$  contains a power of  $\epsilon$  not appearing in the representation of any other. The value of this feasible vector  $\lambda'$  is

$$(\mathbf{c}, \lambda(\epsilon)) = f[\lambda(\epsilon)] = \sum_{i \in I} c_i \lambda_i + \sum_{j=1}^n \epsilon^j z_j,$$

where the  $z_j$  are defined the same as before. Repeating the procedure of (6.1.5) for the modified problem, we have

$$\mathbf{p}_0(\epsilon) = \sum_{i \in I} \left( \lambda_i + \sum_{j=1}^n \epsilon^j x_{ij} - \theta x_{ik} \right) \mathbf{p}_i + \theta \mathbf{p}_k,$$

from which we get a new feasible vector  $\lambda'(\epsilon)$  with value

$$f[\lambda'(\epsilon)] = f[\lambda(\epsilon)] + \theta(c_k - z_k).$$

There are three mutually exclusive possibilities to be considered, as in paragraph D:

- (a) If  $z_k - c_k \geq 0$  for all  $k$ , then  $\lambda$  is maximal.
- (b) If  $z_k - c_k < 0$  for some  $k$ , and  $x_{ik} \leq 0$  for all  $i$  in  $I$ , then  $f(\lambda') \rightarrow \infty$  as  $\theta \rightarrow \infty$  for every  $\epsilon$  and the problem has no solution.
- (c) If  $z_{k_0} - c_{k_0} < 0$  for some  $k_0$ , and  $x_{ik_0} > 0$  for some  $i$  in  $I$ , then let

$$\theta(\epsilon) = \min_i \frac{(\lambda_i + \sum_{j=1}^n \epsilon^j x_{ij})}{x_{ik_0}},$$

the minimum being taken over all  $i$  such that  $x_{ik_0} > 0$ . In practice, as mentioned earlier, it is customary to select the value of  $k_0$  for which  $z_{k_0} - c_{k_0}$  is most negative. Whenever

$$\min_i \frac{\lambda_i}{x_{ik_0}} \quad \begin{matrix} x_{ik_0} > 0 \end{matrix}$$

is achieved at a single index value  $i_0$ , then the same is true for  $\theta(\epsilon)$  with  $\epsilon$  sufficiently small. If  $\min_i (\lambda_i/x_{ik_0})$  is achieved at two or more indices—call the set of such indices  $J$ —then we select the index value of  $J$  for which  $x_{i_1}/x_{ik_0}$  is smallest (algebraically). If this minimum is attained for at least two indices of  $J$ , we proceed to examine the ratio  $x_{i_2}/x_{ik}, x_{i_3}/x_{ik}, \dots$ , determining their minimum over the successively smaller index sets at which the preceding minima were attained, and so on until we find a minimum that is achieved at a single index value  $i_0$ . The process must terminate, since the  $\lambda'_i(\epsilon)$  are all distinct polynomials. The vector with index  $k_0$  replaces the vector with index  $i_0$ , and the algorithm may be repeated.

It is never necessary to calculate  $\lambda^1(\epsilon)$ , since all that is relevant for the next step of the algorithm is the components of  $\lambda_i$  for the new basis of the unmodified problem. The reader may observe that the nondegeneracy Assumption 6.1.1 applies to every modified problem with  $\epsilon$  positive and sufficiently small, and hence that the modified procedure yields ultimately a single solution valid for every such problem. If we then put  $\epsilon = 0$ , this vector also provides a solution to the original problem.

**\*6.2 Auxiliary simplex methods.** In this section we discuss several consequences and modifications of the simplex method.

*A. Solution of the dual problem.* We show first how one can obtain a solution to Problem II, the dual problem, simultaneously with the solution to Problem I. This should help to clarify the dual simplex algorithm which is discussed in paragraph B below.

The simplex method solves the problem of finding a vector  $\lambda$  that maximizes  $(\mathbf{c}, \lambda)$  subject to the constraints  $\mathbf{A}\lambda = \mathbf{b}$  and  $\lambda \geq \mathbf{0}$ . The dual problem is to find a vector  $\mathbf{y}$  that minimizes  $(\mathbf{y}, \mathbf{b})$  subject to the constraints  $\mathbf{y}\mathbf{A} \geq \mathbf{c}$ . This formulation is no less valid when  $\mathbf{A}$  represents an augmented matrix obtained by adding suitable slack variables and defining the corresponding  $c$ 's to be zero, since in that case the constraints  $\mathbf{y}\mathbf{A} \geq \mathbf{c}$  already express the necessary restrictions of the form  $y_i \geq 0$ .

At the end of the simplex algorithm, we arrive at a basis  $\mathbf{B}$ , consisting of a submatrix of  $\mathbf{A}$  whose columns comprise the vectors of the basis, such that

$$\mathbf{B}\lambda^0 = \mathbf{b}, \quad \lambda^0 = \mathbf{B}^{-1}\mathbf{b},$$

where  $\lambda^0$  is the solution of Problem I and the vector  $\mathbf{b}$  is the same as  $\mathbf{p}_0$ . Let  $\mathbf{c}_B$  denote a contraction of the vector  $\mathbf{c}$  restricted to the components corresponding to the basis vectors of  $\mathbf{B}$ . Define  $\mathbf{y}^0 = \mathbf{c}_B \mathbf{B}^{-1}$ . Writing  $\mathbf{A}$  in the form  $\|\mathbf{B}, \mathbf{P}\|$ , where  $\mathbf{P}$  is the set of column vectors of  $\mathbf{A}$  not included in  $\mathbf{B}$ , we have

$$\mathbf{y}^0 \mathbf{A} = \mathbf{y}^0 \|\mathbf{B}, \mathbf{P}\| = \langle \mathbf{c}_B, \mathbf{c}_B \mathbf{B}^{-1} \mathbf{P} \rangle. \quad (6.2.1)$$

Equation (6.2.1) states that  $\mathbf{y}^0 \mathbf{A}$  is the vector whose first  $m$  components are those of  $\mathbf{c}_B$  and whose last  $n - m$  components are those of  $\mathbf{c}_B \mathbf{B}^{-1} \mathbf{P}$ . We also note that the tableau associated with the basis  $\mathbf{B}$ , when the vectors of  $\mathbf{B}$  are arranged first, is clearly  $\|\mathbf{I}, \mathbf{B}^{-1} \mathbf{P}\|$ . Hence the vector  $\langle \mathbf{c}_B, \mathbf{c}_B \mathbf{B}^{-1} \mathbf{P} \rangle$  is identical to the vector  $\mathbf{z}$  (see page 165) for the basis  $\mathbf{B}$ ; and since  $\mathbf{B}$  is optimal,  $\mathbf{z} \geq \mathbf{c}$ . Thus  $\mathbf{y}^0$  is feasible for the dual problem. Moreover,

$$(\mathbf{y}^0, \mathbf{b}) = (\mathbf{c}_B \mathbf{B}^{-1}, \mathbf{b}) = (\mathbf{c}_B, \mathbf{B}^{-1} \mathbf{b}) = (\mathbf{c}, \lambda^0). \quad (6.2.2)$$

It follows from Corollary 5.4.1 that  $\mathbf{y}^0$  is optimal for the dual problem.

*B. The dual simplex algorithm.* We shall refer to the simplex method of Section 6.1 hereafter as the primal simplex algorithm, to distinguish it from the dual simplex algorithm described below. The dual simplex algorithm furnishes another algorithm for solving Problem I. It is referred to as the dual algorithm, since its underlying justification is bound intrinsically to the solution of the dual problem. We shall discuss briefly the merits and defects of each of these algorithms at the close of this section.

A solution  $\lambda^0$  to Problem I is determined whenever a matrix tableau  $\mathbf{X}$  is achieved for a basis satisfying the property that  $z_j - c_j \geq 0$  for all  $j$ , where

$$z_k = \sum_{j \in B} c_j x_{jk},$$

the summation extending over the indices corresponding to the basis vectors. The main idea of the transformations underlying a change of basis is to eliminate step by step the negative components of  $z_j - c_j$  while maintaining  $\lambda$  as a feasible vector.

A dual method of calculation is to insist that the vector  $\mathbf{z} - \mathbf{c}$  always be nonnegative but permit the vector  $\lambda$  used in the representation of  $\mathbf{p}_0$  to possess components of arbitrary sign. This is equivalent to maintaining feasibility for the dual problem and proceeding until a feasible vector for the original problem is obtained. A new basis is formed by replacing a vector corresponding to a negative coordinate of  $\lambda$  by a new vector. The process is repeated and a solution is achieved after a finite number of steps.

Since we are concerned here only with the salient features of the dual algorithm, and not with its many special cases, we shall impose the following nondegeneracy assumption.

ASSUMPTION 6.2.1. For no feasible  $\mathbf{y}$  of the dual problem can equality of  $\mathbf{y}\mathbf{A}$  and  $\mathbf{c}$  be achieved at more than  $m$  coordinates.

The formal procedure is as follows. Suppose that we have a basis of column vectors  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m$  with matrix  $\mathbf{Q}$  such that  $\mathbf{Q}\mathbf{u} = \mathbf{b}$ ; the components of  $\mathbf{u}$  are not necessarily all nonnegative. Let  $\mathbf{X}$  represent the tableau associated with  $\mathbf{Q}$ , and let  $\mathbf{c}_Q$  denote the vector  $\mathbf{c}$  contracted to the components corresponding to the columns of  $\mathbf{Q}$ . We further assume that  $z_j - c_j \geq 0$  for all  $j$ , with

$$z_j = \sum_{i \in Q} c_i x_{ij}.$$

As we see from paragraph A of this section, this is equivalent to saying that  $\mathbf{y} = \mathbf{c}_Q \mathbf{Q}^{-1}$  is a feasible vector for the dual problem. Because of Assumption 6.2.1, equality holds for  $\mathbf{y}\mathbf{A} \geq \mathbf{c}$  at the components corresponding to the basis vectors, with strict inequality elsewhere.

If  $u_i \geq 0$  for every component, then  $\mathbf{u}$  is optimal by the primal simplex algorithm.

Let  $r_0$  denote the index for which  $u_i$  is most negative. We now distinguish two cases.

*Case 1.* If  $x_{r_0j} \geq 0$  for all  $j$ , then there exists no feasible vector satisfying the constraints of Problem I.

*Proof.* Analogous to (6.1.5), we have

$$\mathbf{p}_0 = \sum_{l \in Q} (u_l - \theta x_{lk}) \mathbf{q}_l + \theta \mathbf{q}_k, \quad (6.2.3)$$

where  $k$  is arbitrary and  $\theta$  is any real number. Suppose that there exists a representation

$$\mathbf{p}_0 = \sum_{j=1}^m \lambda_j \mathbf{q}_{kj},$$

with  $\lambda_j \geq 0$ . Inserting  $\theta = \lambda_j$  and  $k = k_j$  successively into (6.2.3) and adding, we obtain

$$m\mathbf{p}_0 = \sum_{l \in Q} \left[ mu_l - \sum_j \lambda_j x_{lk_j} \right] \mathbf{q}_l + \mathbf{p}_0$$

or

$$\mathbf{p}_0 = \sum_{l \in Q} \left[ \frac{m}{m-1} u_l - \frac{\sum_j \lambda_j x_{lk_j}}{m-1} \right] \mathbf{q}_l. \quad (6.2.4)$$

Since the vectors  $\mathbf{q}_l$  constitute a basis, the coefficient of  $\mathbf{q}_l$  in (6.2.4) is equal to  $u_l$ , and in particular

$$u_{r_0} = \frac{m}{m-1} u_{r_0} - \frac{\sum_j \lambda_j x_{r_0 k_j}}{m-1}$$

or

$$0 = \frac{1}{m-1} u_{r_0} - \frac{\sum_j \lambda_j x_{r_0 k_j}}{m-1}.$$

But  $u_{r_0} < 0$ ,  $\lambda_j \geq 0$ , and  $x_{r_0 k} \geq 0$  (all  $k$ ) contradict this relation. Therefore there exists no feasible primal vector. Since the dual problem has feasible vectors  $(c_Q Q^{-1})$ , this means that the dual form tends to  $-\infty$ .

*Case 2.* Let  $u_{r_0} < 0$  and  $x_{r_0 j} < 0$  for some  $j$ . (Necessarily,  $j$  is the index of a vector outside the basis.)

Define  $k_0$  as the index value at which

$$\alpha = \min_j \left( \frac{z_j - c_j}{-x_{r_0 j}} \right) \quad x_{r_0 j} < 0$$

is achieved. From the nondegeneracy assumption it readily follows that  $k_0$  is unique. We now choose  $\theta$  in (6.2.3) so that the coefficient of  $q_{r_0}$  vanishes, viz.,

$$\theta = \frac{u_{r_0}}{x_{r_0 k_0}}.$$

Since both  $u_{r_0}$  and  $x_{r_0 k_0}$  are negative, we have  $\theta$  positive. A new basis  $\tilde{Q}$  is obtained from the old by replacing  $q_{r_0}$  by  $q_{k_0}$ . The coefficients in the representation of  $p_0$  are

$$\tilde{u}_l = u_l - \theta x_{l k_0} \quad \text{and} \quad \tilde{u}_{k_0} = \theta. \quad (6.2.5)$$

A straightforward computation, utilizing the transformation of basis (6.1.8) and the definition of  $\alpha$ , shows that  $\tilde{z} - \tilde{c}$  for the new basis  $\tilde{Q}$  is a nonnegative vector and hence that  $c_{\tilde{Q}} \tilde{Q}^{-1}$  is a feasible vector for the dual. On comparison with (6.2.2), we get

$$(c_{\tilde{Q}} \tilde{Q}^{-1}, b) = \sum_{l \in \tilde{Q}} \tilde{u}_l c_l = \sum_{l \in Q} u_l c_l - \theta z_{k_0} + \theta c_{k_0}$$

and

$$(c_Q Q^{-1}, b) = \sum_l u_l c_l.$$

Therefore

$$(c_Q Q^{-1}, b) - (c_{\tilde{Q}} \tilde{Q}^{-1}, b) = \theta(z_{k_0} - c_{k_0}) > 0,$$

since  $\theta > 0$  and  $z_{k_0} - c_{k_0} > 0$ ; it follows from Assumption 6.2.1 that  $c_{\tilde{Q}} \tilde{Q}^{-1}$  is an improvement over  $c_Q Q^{-1}$  for the dual problem. Both vectors are extreme and feasible, and therefore the procedure terminates with  $u$  nonnegative in a finite number of steps.

The merit of the dual simplex method is that it involves far fewer

iterations than the usual simplex method for certain kinds of problems. For example, consider the problem

$$\min (\mathbf{c}, \mathbf{y}) \quad (6.2.6)$$

subject to the constraints

$$\mathbf{y} \geq \mathbf{0}, \quad \mathbf{yA} \geq \mathbf{b}. \quad (6.2.7)$$

This problem can be converted into a maximization problem by substituting  $-\mathbf{A}$  for  $\mathbf{A}$ ,  $-\mathbf{b}$  for  $\mathbf{b}$ , and  $-\mathbf{c}$  for  $\mathbf{c}$ . With this done, we can formulate the problem as

$$\max (\mathbf{c}, \mathbf{y})$$

subject to the constraints

$$\mathbf{y} \geq \mathbf{0}, \quad \mathbf{yA} \leq \mathbf{b}.$$

Now if  $\mathbf{c}$  and  $\mathbf{b}$  of (6.2.6) and (6.2.7) are positive vectors, then in the changeover to a maximization problem they become negative vectors. In this case, in order to get started for the primal simplex algorithm, it seems necessary to introduce artificial variables on top of the slack variables, thus markedly increasing the number of iterations. In these circumstances, it appears that the dual method would yield a solution more quickly.

Indeed, suppose that the tableau  $\mathbf{T}$  associated with the vectors of the slack variables is

$$\mathbf{T} = \|\mathbf{A}, \mathbf{I}\|.$$

Since the  $c_i$  attached to the slack variables are defined to be zero, we get for this basis

$$z_i - c_i = -c_i,$$

which is a positive vector if all  $c_i \leq 0$ . Hence, the dual method can be started immediately with an initial basis whose tableau is  $\mathbf{T}$ . The use of artificial variables in this circumstance would be entirely redundant.

The classical diet problem (see Section 5.1) is of the sort described in (6.2.6) and (6.2.7). It is more readily solved by the dual algorithm than by the primal.

Many variations of the simplex method have been introduced to take advantage of special forms of the constraint matrix. For example, the problem created by adding a single additional constraint to a problem already solved by the simplex method can be solved without having to repeat the full procedure. Another special case, the problem in which the variables are restricted by upper bounds in addition to the usual constraints, can be solved without doubling the number of variables. Descriptions of some of these special techniques are given in several elementary texts in linear programming (see notes to this chapter).

**6.3 An illustration of the use of the simplex method.** A manufacturer has the option of using one or more of four different production processes. The first and second processes yield items which we call  $A$ , and the third and fourth yield items  $B$ . The inputs for each of these processes are labor measured in man-weeks, pounds of raw material  $w$ , and boxes of raw material  $\omega$ . Since each process has different input requirements, the profits yielded by the various processes differ, even for processes producing the same item. The manufacturer, in deciding on a week's production schedule, is limited in the range of possibilities by the available amounts of the two raw materials and of manpower. Table 6.1 describes the situation.

TABLE 6.1

	One item of $A$		One item of $B$		Availabilities
	Process 1	Process 2	Process 3	Process 4	
Man-weeks	1	1	1	1	15
Pounds of material $w$	7	5	3	2	120
Boxes of material $\omega$	3	5	10	15	100
Unit profit \$	6	5.5	9	8	
Production levels	$x_1$	$x_2$	$x_3$	$x_4$	

Every production plan  $\langle x_1, x_2, x_3, x_4 \rangle$  must obey the availability restrictions

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &\leq 15, \\ 7x_1 + 5x_2 + 3x_3 + 2x_4 &\leq 120, \\ 3x_1 + 5x_2 + 10x_3 + 15x_4 &\leq 100. \end{aligned} \quad (6.3.1)$$

Furthermore, we do not allow "negative production"; i.e., we require

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0. \quad (6.3.2)$$

Our problem is to find a set of values for the unknown  $x$ 's which satisfy the relations (6.3.1) and (6.3.2), and which maximize total profit  $r_0$ , where

$$r_0 = 6x_1 + 5.5x_2 + 9x_3 + 8x_4. \quad (6.3.3)$$

If we explicitly include unused materials or idle man-weeks as variables—

$$\begin{aligned} x_5 &= \text{slack (unused) man-weeks,} \\ x_6 &= \text{slack (unused) material } w, \\ x_7 &= \text{slack (unused) material } \omega \end{aligned} \quad (6.3.4)$$

—where again

$$x_5 \geq 0, \quad x_6 \geq 0, \quad x_7 \geq 0, \quad (6.3.5)$$

then we can convert the relations in (6.3.1) to equalities:

$$x_1 + x_2 + x_3 + x_4 + x_5 = 15,$$

$$7x_1 + 5x_2 + 3x_3 + 2x_4 + x_6 = 120,$$

$$3x_1 + 5x_2 + 10x_3 + 15x_4 + x_7 = 100,$$

without altering the profit function (6.3.3).

The various simplex iterations are described in Tables 6.2, 6.3, and 6.4, the starting basis (Table 6.2) consisting of the vectors corresponding to the slack variables. Note that the components of  $\mathbf{z}$  are the dot products of the columns of the “tableau” with the column  $\mathbf{c}$  on the left. The components of  $\boldsymbol{\lambda}$  appear under  $\mathbf{p}_0$ .

TABLE 6.2

$\mathbf{c}$	6	5.5	9	8	0	0	0	
Basis	$\mathbf{p}_0$	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{p}_4$	$\mathbf{p}_5$	$\mathbf{p}_6$	$\mathbf{p}_7$
0 $\mathbf{p}_5$	15	1	1	1	1	1	0	0
0 $\mathbf{p}_6$	120	7	5	3	2	0	1	0
0 $\mathbf{p}_7$	100	3	5	10	15	0	0	1
$\mathbf{z}$	0	0	0	0	0	0	0	0
$\mathbf{z} - \mathbf{c}$	-6	-5.5	-9	-8	0	0	0	0

The largest negative component of  $\mathbf{z} - \mathbf{c}$  is  $-9$  and the smallest ratio of  $\lambda_i^0/x_{ik_0}$  is  $100/10$ . Hence vector  $\mathbf{p}_3$  replaces  $\mathbf{p}_7$  for the next basis (Table 6.3).

TABLE 6.3

$\mathbf{c}$ Basis	$\mathbf{p}_0$	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{p}_4$	$\mathbf{p}_5$	$\mathbf{p}_6$	$\mathbf{p}_7$
0 $\mathbf{p}_5$	5	0.7	0.5	0	-0.5	1	0	-0.1
0 $\mathbf{p}_6$	90	6.1	3.5	0	-2.5	0	1	-0.3
9 $\mathbf{p}_3$	10	0.3	0.5	1	1.5	0	0	0.1
$\mathbf{z}$		2.7	4.5	9	13.5	0	0	0.9
$\mathbf{z} - \mathbf{c}$		-3.3	-1	0	5.5	0	0	0.9

Here the largest negative component of  $\mathbf{z} - \mathbf{c}$  is  $-3.3$  and the smallest ratio is  $5/0.7$ . Hence vector  $\mathbf{p}_1$  replaces  $\mathbf{p}_5$ .

TABLE 6.4

c	Basis	p <sub>0</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>5</sub>	p <sub>6</sub>	p <sub>7</sub>
6	p <sub>1</sub>	7.143	1	0.714	0	-0.714	1.429	0	-0.143
0	p <sub>6</sub>	46.868	0	-0.855	0	1.855	8.717	1	0.572
9	p <sub>3</sub>	7.857	0	0.288	1	1.714	-0.429	0	0.143
	<b>z</b>		6	6.886	9	11.142	4.713	0	0.429
	<b>z - c</b>		0	1.386	0	3.142	4.713	0	0.429

Since  $z_i - c_i$  is nonnegative for all  $i$  (Table 6.4), we have secured a solution:

$$x_1^0 = 7.143, \quad x_3^0 = 7.857, \quad x_6^0 = 46.868.$$

The optimal production plan consistent with the constraints calls for 7.143 units of item  $A$  using process 1, and 7.857 units of item  $B$  using process 3. The total profit earned under this plan is

$$(7.143)6 + (7.857)9 = 113.571.$$

All available man-weeks and all available material  $w$  are used in this plan, but since  $x_6^0 > 0$ , some pounds of material  $w$  are not used. In this example it can be shown that the solution is unique.

**\*6.4 Computation of network flow.** In this section we describe a special algorithm for the solution of maximal network flow problems (see Sections 5.9 and 5.10) in which there are capacity constraints on arcs only.

Consider a network consisting of nodes  $p_i$  ( $i = 1, \dots, n$ ) and arcs  $p_i p_j$  (not necessarily directed) joining certain pairs of nodes. Let  $p_1$  be the source,  $p_n$  the sink, and  $p_i$  ( $i = 2, \dots, n-1$ ) intermediate nodes. Associated with each arc  $p_i p_j$  is a nonnegative capacity  $c_{ij}$  representing an upper bound on flow from  $p_i$  to  $p_j$ . (Throughout what follows we have assumed  $c_{ii} = 0$ .) If  $x_{ij}$  is the flow from  $p_i$  to  $p_j$ , and  $v$  denotes the net flow leaving  $p_1$  (the flow value), the problem may be stated as follows: Maximize  $v$  subject to

$$\begin{aligned} \sum_j (x_{1j} - x_{j1}) &= v, \\ \sum_j (x_{ij} - x_{ji}) &= 0 \quad (i = 2, \dots, n-1), \\ \sum_j (x_{nj} - x_{jn}) &= -v, \end{aligned} \tag{6.4.1}$$

and

$$0 \leq x_{ij} \leq c_{ij}. \tag{6.4.2}$$

The constraints (6.4.1) state that the net flow leaving the source (or entering the sink) is  $v$ , and that the flow into any intermediate node is equal to the flow out; (6.4.2) states that the flow in any arc does not exceed its capacity. Arcs not present in the network may be assumed to have zero capacity, or alternatively we may think of the corresponding variables  $x_{ij}$  as suppressed.

This is, of course, a linear programming problem, which means that it may be solved by the simplex method. In fact, it can be shown that the simplex computation for this problem always involves solving triangular systems of equations, and hence is very efficient. In this section, however, we present an even simpler and faster algorithm for the solution of maximal flow problems. Like the simplex method, it yields solutions to both the primal and dual problems.

Before describing the computation in general terms, let us consider a simple example. Figure 6.1 is a network with source  $p_1$ , sink  $p_4$ , and arc capacities as indicated. We ask whether the flow shown in Fig. 6.2, namely,  $x_{13} = x_{32} = x_{24} = 1$ ,  $x_{12} = x_{23} = x_{34} = 0$  (with value  $v = 1$ ), is maximal. (In each arc the first number is the capacity and the second is the actual flow.) The answer is no, since we may, for example, increase  $x_{12}$ ,  $x_{23}$ , and  $x_{34}$  each by one unit, yielding the flow shown in Fig. 6.3, which has value  $v = 2$ .

Is this new flow maximal? Again the answer is no, since we may increase  $x_{12}$ , decrease  $x_{32}$ , and increase  $x_{34}$ , each by one unit, obtaining the flow shown in Fig. 6.4, which has value  $v = 3$ .

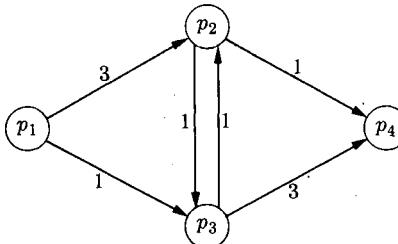


FIGURE 6.1

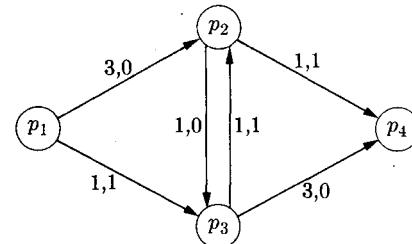


FIGURE 6.2

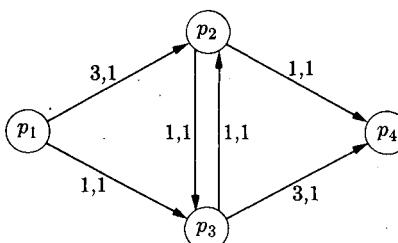


FIGURE 6.3

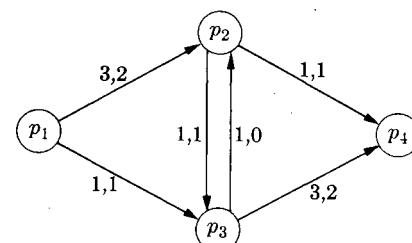


FIGURE 6.4

The flow is now maximal, since the cut consisting of arcs  $p_1p_3$ ,  $p_2p_3$ , and  $p_2p_4$  also has value  $v = 3$ .

Notice that each of the flow improvements in the example was accomplished by altering a sequence of variables  $x_{ij}$  along a "path" from source to sink, by either adding or subtracting a positive constant from each arc flow along the path. The algorithm which we now describe presents a systematic and efficient method of searching for flow improvements of this kind. If no such improvement exists, it will be shown that a maximal flow has been attained.

To ensure termination of the computation, we make the assumption that all the arc capacities are nonnegative integers. Since a problem with rational  $c_{ij}$  can be converted to one with integral  $c_{ij}$  simply by clearing fractions, this is not, computationally speaking, a drastic restriction.

The computation may start with any integral flow  $\mathbf{x}$  (for example, the zero flow). For some of the nodes  $p_i$ , we define labels recursively as follows:

Step (a): Start by considering  $p_1$  as labeled and unscanned.

Step (b): Take any labeled, unscanned node  $p_i$ , and scan  $p_i$  for all unlabeled nodes  $p_j$  such that  $x_{ij} < c_{ij}$ . Label such nodes  $p_j^+$ . Next scan  $p_i$  for all unlabeled nodes  $p_j$  such that  $x_{ji} > 0$ . Label these nodes  $p_j^-$ . Consider  $p_i$  as scanned.

Step (b) is repeated either until the sink  $p_n$  is labeled, or until no further labeling is possible and  $p_n$  is unlabeled. In the latter case, the computation terminates; in the former case, the flow  $\mathbf{x}$  may be improved along a path from  $p_1$  to  $p_n$  as follows. If  $p_n$  has been labeled  $p_j^+$ , tentatively add  $\epsilon > 0$  to  $x_{jn}$ ; if  $p_n$  has been labeled  $p_j^-$ , subtract  $\epsilon$  from  $x_{nj}$ . Next look at the label on  $p_j$ . If this is  $p_i^+$ , add  $\epsilon$  to  $x_{ij}$  (if  $p_i^-$ , subtract  $\epsilon$  from  $x_{ji}$ ) and proceed to  $p_i$  and its label. Continue in this fashion until the source has been reached. Since for any  $\epsilon$ , the constraints (6.4.1) remain satisfied with a new flow value  $v + \epsilon$ , we take  $\epsilon$  as large as possible consistent with (6.4.2). Clearly  $\epsilon$ , being the minimum of certain positive integers, is itself a positive integer. Thus we arrive at a new feasible flow vector with a flow value in excess of the preceding by at least one unit.

The algorithm is then repeated with this new flow as a starting point and we obtain a new labeled process. If  $p_n$  is labeled, we may again construct a feasible flow vector with total flow value at least one unit more. Obviously, after a finite number of iterations the procedure must terminate with a flow whose corresponding labeled process does not include  $p_n$  as a labeled node.

It remains to show that if the labeling process ceases with  $p_n$  unlabeled, the flow  $\mathbf{x}$  is maximal. Let  $S$  be the set of indices  $i$  for which  $p_i$  is labeled. Thus  $1 \in S$ ,  $n \notin S$ . Now sum equations (6.4.1) over  $i \in S$ , obtaining

$$v = \sum_{\substack{i \in S \\ j \notin S}} (x_{ij} - x_{ji}),$$

the terms  $j \in S$  having canceled out in summation. But  $i \in S, j \notin S$  imply  $x_{ij} = c_{ij}$  and  $x_{ji} = 0$ , for otherwise  $p_j$  would have received a label. Hence

$$v = \sum_{\substack{i \in S \\ j \notin S}} c_{ij}.$$

The set of arcs leading from  $S$  to its complement is a cut; otherwise there must exist a flow path from  $p_1$  to  $p_n$  all of whose nodes belong to  $S$ , from which it follows that  $p_n$  is labeled, contrary to hypothesis. By Theorem 5.10.1 the flow  $\mathbf{x}$  is maximal and the cut is minimal.

We conclude this section by indicating a type of problem which has constraining equations very similar to those of the transportation problem, but which, in fact, can also be identified as a network flow problem: viz., find  $\mathbf{X} = \|x_{ij}\|$  which maximizes  $\sum_{i,j} x_{ij}$  where

$$\sum_{j=1}^m x_{ij} \leq a_i, \quad \sum_{i=1}^n x_{ij} \leq b_j, \quad (6.4.3)$$

and  $x_{ij} = 0$  for a prescribed set  $\Omega$  of pairs  $(i, j)$ . To see this as a flow problem, set up the directed network of  $m + n + 2$  nodes as shown in Fig. 6.5.

The capacity on the directed arc  $P_i Q_j$  is zero if  $i, j \in \Omega$ , and large otherwise; the capacities of the source and sink arcs are the  $a_i$  and  $b_j$  as shown, and  $x_{ij}$  is interpreted as the flow from  $P_i$  to  $Q_j$ .

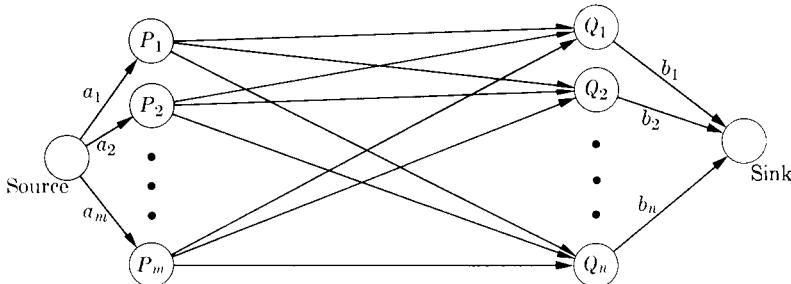


FIGURE 6.5

**6.5 A method of approximating the value of a game.** The computational methods discussed in the previous sections dealt directly with linear programming problems. The method described in this section provides a means of approximating the value of a game. This method enables us to construct a sequence of strategies for each player such that all limit strategies secured from this sequence are optimal strategies for the game. In particular, if both players have unique optimal strategies, these optimal

strategies may be approximated with any prescribed degree of accuracy. If a linear programming problem and its dual problem have unique solutions, then by considering the equivalent game problem we can obtain another iterative scheme for solving the problem, and vice versa.

The method of this section is founded on the following intuitive considerations. Suppose that the game is played repeatedly, and that on each successive playing each player chooses a strategy on the assumption that his opponent's future actions will resemble the past. If what each player did on the first playing is known, this principle for selecting a strategy leads to a determinate sequence of plays, except for a slight arbitrariness which arises if several alternatives happen at an intermediate play (in this case any choice among the alternatives can be made without invalidating the convergence properties established below). From this sequence one can calculate upper and lower bounds for the value of the game and approximate the optimal strategy for each player.

The formal description of the iterations follows.

Let  $\mathbf{A} = \{a_{ij}\}$  be the pay-off matrix. Let  $\mathbf{a}_i$  denote the  $i$ th row and  $\mathbf{b}_j$  the  $j$ th column of  $\mathbf{A}$ .

Assume that for the first play of the game Player II chooses a pure strategy corresponding to some column of  $\mathbf{A}$  which we shall denote by  $\mathbf{c}(1)$ . In order to maximize his gain, Player I must then choose the pure strategy corresponding to the row index for which the maximum of  $\mathbf{c}(1)$  is achieved. We denote this row of  $\mathbf{A}$  by  $\mathbf{r}(1)$ . Thus, if  $\mathbf{c}(1)$  and  $\mathbf{r}(1)$  are the  $j_0$ th column and the  $i_0$ th row, respectively, we have

$$a_{i_0 j_0} = \max_k a_{k j_0}, \quad \text{or} \quad c_{i_0}(1) = \max_k c_k(1).$$

Turning to the second play of the game, if Player I chooses the strategy whose yields are the components of the row  $\mathbf{r}(1)$ , Player II might counter with the pure strategy  $j_1$ , where

$$a_{i_0 j_1} = \min_j a_{i_0 j} = \min_j r_j(1).$$

However, for continuity we suppose that both players modify their past choice of a strategy only in the light of their experience. Actually, each plays the pure strategy that is optimal in terms of his total experience with his opponent, assigning equal weights to his opponent's previous actions.

Specifically, suppose that Player II chooses the mixed strategy  $\langle \frac{1}{2}, \frac{1}{2} \rangle$  on  $j_0$  and  $j_1$ .

Let  $\mathbf{c}(2) = \mathbf{c}(1) + \mathbf{b}_{j_1}$ .

Now Player I tries to maximize his gain again. He proceeds to determine  $\max_i \frac{1}{2}[c_i(1) + b_{i j_1}]$ , or equivalently  $\max_i c_i(2)$ . If this maximum occurs at  $i_1$ , he plays the average of  $\mathbf{r}(1)$  and  $\mathbf{a}_{i_1}$ .

Let  $\mathbf{r}(2) = \mathbf{r}(1) + \mathbf{a}_{i_1}$ .

We continue this fictitious style of play in which each player in turn responds optimally to his opponent's previous plays. Next Player II finds a column index  $j_2$  at which  $\min_j r_j(2)$  is achieved and plays the mixed strategy  $\langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$  on  $j_0, j_1, j_2$ .

Let  $\mathbf{c}(3) = \mathbf{c}(2) + \mathbf{b}_{j_2}$ .

In answer to this strategy, Player I selects the largest component of  $\frac{1}{3}\mathbf{c}(3)$ . If the maximum occurs at  $i_2$ , he plays  $\langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$  on  $i_0, i_1, i_2$ , and so on. We thus obtain two sequences of vectors:

$$\mathbf{c}(k) = \mathbf{c}(k-1) + \mathbf{b}_{j_{k-1}}, \quad \text{where} \quad r_{j_{k-1}}(k-1) = \min_j r_j(k-1),$$

and

$$\mathbf{r}(k) = \mathbf{r}(k-1) + \mathbf{a}_{i_{k-1}}, \quad \text{where} \quad c_{i_k}(k) = \max_i c_i(k).$$

We shall demonstrate in the next section that the maximum and minimum pay-offs corresponding to the strategies  $\mathbf{c}(n)/n, \mathbf{r}(n)/n$  converge to  $v$ , i.e., that

$$\lim_n \left[ \max_i \frac{c_i(n)}{n} \right] = \lim_n \left[ \min_j \frac{r_j(n)}{n} \right] = v.$$

First, however, we shall consider briefly the analogous computing scheme formulated directly in terms of mixed strategies.

We assume that Player II starts with some mixed strategy  $\mathbf{y}_1$ . Player I then chooses  $\mathbf{x}_1$ , where

$$\max_x (\mathbf{x}, \mathbf{A}\mathbf{y}_1) = (\mathbf{x}_1, \mathbf{A}\mathbf{y}_1).$$

Player II next chooses  $\mathbf{y}_2$ , where

$$\min_y (\mathbf{x}_1, \mathbf{A}\mathbf{y}) = (\mathbf{x}_1, \mathbf{A}\mathbf{y}_2).$$

Assume that Player II plays the average  $\frac{1}{2}(\mathbf{y}_1 + \mathbf{y}_2)$  on this second round; Player I then selects  $\mathbf{x}_2$ , where

$$\max_x \left( \mathbf{x}, \mathbf{A} \left( \frac{\mathbf{y}_1 + \mathbf{y}_2}{2} \right) \right) = \left( \mathbf{x}_2, \mathbf{A} \left( \frac{\mathbf{y}_1 + \mathbf{y}_2}{2} \right) \right).$$

Each player continues to play the average of all his choices, and the opposing player tries to improve his chances against this average. For instance, if  $\mathbf{x}_n$  is Player I's best choice against

$$\frac{1}{n} [\mathbf{y}_1 + \cdots + \mathbf{y}_n],$$

he will choose the strategy

$$\frac{1}{n} [\mathbf{x}_1 + \cdots + \mathbf{x}_n];$$

then Player II will play

$$\frac{1}{n+1} [\mathbf{y}_1 + \cdots + \mathbf{y}_{n+1}],$$

etc. It can be established that

$$\begin{aligned} \lim_n \left[ \max_x \left( \mathbf{x}, \frac{1}{n} \mathbf{A}(\mathbf{y}_1 + \cdots + \mathbf{y}_n) \right) \right] \\ = \lim_n \left[ \min_y \left( \frac{1}{n} (\mathbf{x}_1 + \cdots + \mathbf{x}_n), \mathbf{A}\mathbf{y} \right) \right] = v, \quad (6.5.1) \end{aligned}$$

and that any limit point of

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{or} \quad \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$$

is an optimal strategy for the game. We shall prove the convergence of the scheme dealing with columns and rows, but a modification of our proof yields (6.5.1).

The following example illustrates the method. Consider the game defined by the pay-off matrix

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 0 \\ 2 & 0 & 3 \\ -1 & 3 & -3 \end{vmatrix}.$$

Suppose that Player I begins with  $i_0$ , Player II selects  $j_0$ , Player I comes back with  $i_1$ , etc., each player making alternate choices according to the scheme described above. Let

$$\underline{v}(N) = \frac{1}{N} \min_j r_j(N), \quad \bar{v}(N) = \frac{1}{N} \max_i c_i(N).$$

The figures are tabulated in Table 6.5. The value of the game is 1.

**\*6.6 Proof of the convergence.** A sequence of vectors  $\mathbf{c}(k)$  and  $\mathbf{r}(k)$ , of dimension  $n$  and  $m$ , respectively, is said to form a “vector system” with respect to an  $n \times m$  matrix  $\mathbf{A}$  if

- (1)  $\max_i c_i(0) = \min_j r_j(0)$ ;
- (2) where  $\max_i c_i(k)$  is achieved for the  $v$ th component,

$$\mathbf{r}(k+1) = \mathbf{r}(k) + \mathbf{a}_v;$$

TABLE 6.5

$N$	$i_N$	$r_1(N)$	$r_2(N)$	$r_3(N)$	$\underline{v}(N)$	$j_N$	$c_1(N)$	$c_2(N)$	$c_3(N)$	$\bar{v}(N)$
1	1	2	1	0	0.000	3	0	3	-3	3.000
2	2	4	1	3	0.500	2	1	3	0	1.500
3	2	6	1	6	0.333	2	2	3	6	1.000
4	2	8	1	9	0.250	2	3	3	6	1.500
5	3	7	4	6	0.800	2	4	3	9	1.800
6	3	8	7	3	0.500	3	4	6	6	1.000
7	2	8	7	6	0.857	3	4	9	3	1.286
8	2	10	7	9	0.875	2	5	9	6	1.125
9	2	12	7	12	0.778	2	6	9	9	1.000
10	2	14	7	15	0.700	2	7	9	12	1.200
11	3	13	10	12	0.909	2	8	9	15	1.364
12	3	12	13	9	0.750	3	8	12	12	1.000
13	2	14	13	12	0.923	3	8	15	9	1.154
14	2	16	13	15	0.929	2	9	15	12	1.071
15	2	18	13	18	0.867	2	10	15	15	1.000
16	2	20	13	21	0.812	2	11	15	18	1.125
17	3	19	16	18	0.941	2	12	15	21	1.235
18	3	18	19	15	0.833	3	12	18	18	1.000

(3) where  $\min_j r_j(k)$  is achieved for the  $\mu$ th component,

$$\mathbf{c}(k) = \mathbf{c}(k-1) + \mathbf{b}_\mu$$

( $\mathbf{a}_v$  denotes the  $v$ th row and  $\mathbf{b}_\mu$  the  $\mu$ th column of  $\mathbf{A}$ ).

We can readily see that the vectors  $\mathbf{c}(k)$  and  $\mathbf{r}(k)$  of the scheme introduced in Section 6.5 form a vector system if  $\mathbf{c}(0) \equiv \mathbf{r}(0) \equiv \mathbf{0}$ .

► **THEOREM 6.6.1.** For any vector system,

$$\lim_n \left[ \max_i \frac{c_i(n)}{n} \right] = \lim_n \left[ \min_j \frac{r_j(n)}{n} \right] = v.$$

It is conjectured that the process converges at a rate  $1/\sqrt{k}$ , where  $k$  is the number of iterations. In actual cases, it is found that the process is far more efficient than is expected theoretically.

► **LEMMA 6.6.1.** For any vector system,

$$\lim_{T \rightarrow \infty} \left[ \frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} \right] \geq 0.$$

By  $\max \mathbf{c}(T)$  we mean the maximum component of the vector  $\mathbf{c}(T)$ .

*Proof. Special case.*  $\mathbf{c}(0) = \mathbf{r}(0) = \mathbf{0}$ . Then

$$\max \frac{\mathbf{c}(T)}{T} \geq v \geq \min \frac{\mathbf{r}(T)}{T},$$

and this implies the lemma.

*General case.* Let  $n_i, m_j$  denote the number of times that  $\mathbf{a}_i, \mathbf{b}_j$  are added to  $\mathbf{r}(0), \mathbf{c}(0)$  in forming  $\mathbf{r}(T), \mathbf{c}(T)$ , respectively. Then

$$\sum n_i = \sum m_j = T,$$

and

$$x_i = \frac{n_i}{T}, \quad y_j = \frac{m_j}{T}$$

are strategies for the game. We observe that

$$\frac{c_i(T)}{T} = \frac{c_i(0)}{T} + \sum_j a_{ij} \frac{m_j}{T} = \frac{c_i(0)}{T} + (\mathbf{Ay})_i,$$

$$\frac{r_j(T)}{T} = \frac{r_j(0)}{T} + \sum_i a_{ij} \frac{n_i}{T} = \frac{r_j(0)}{T} + (\mathbf{xA})_j,$$

and

$$\frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} = \max_i \frac{\mathbf{c}(0)}{T} - \min_j \frac{\mathbf{r}(0)}{T} + \max_i (\mathbf{Ay})_i - \min_j (\mathbf{xA})_j. \quad (6.6.1)$$

But

$$\lim_T \max_i \frac{\mathbf{c}(0)}{T} = \lim_T \min_j \frac{\mathbf{r}(0)}{T} = 0. \quad (6.6.2)$$

Also, by the min-max theorem and Lemma 1.4.1,

$$v = \min_y \max_x (\mathbf{x}, \mathbf{Ay}) = \min_y \max_i (\mathbf{Ay})_i \leq \max \mathbf{Ay}.$$

Similarly,

$$v \geq \min_j (\mathbf{xA})_j \rightarrow \max \mathbf{Ay} - \min \mathbf{xA} \geq 0. \quad (6.6.3)$$

Relations (6.6.1), (6.6.2), and (6.6.3) prove the lemma.

**DEFINITION.** A row  $\mathbf{a}_\nu$  of  $\mathbf{A}$  is said to be eligible in the interval  $(M, M + T)$  if there exists  $R$  ( $M \leq R \leq M + T$ ) such that

$$\max \mathbf{c}(R) = c_\nu(R).$$

A column  $\mathbf{b}_\mu$  is eligible if

$$\min \mathbf{r}(R) = r_\mu(R).$$

This definition does not imply that  $\mathbf{a}_\nu, \mathbf{b}_\mu$  are actually used in  $\mathbf{r}(R), \mathbf{c}(R)$ .

►LEMMA 6.6.2. If all rows and columns of  $\mathbf{A}$  are eligible in  $(M, M + T)$ , then

$$\max \mathbf{c}(M + T) - \min \mathbf{c}(M + T) \leq 2aT \quad (6.6.4a)$$

and

$$\max \mathbf{r}(M + T) - \min \mathbf{r}(M + T) \leq 2aT, \quad (6.6.4b)$$

where

$$a = \max_{i,j} |a_{ij}|.$$

*Proof.* We shall prove only the first inequality. Let  $\nu$  denote any row for which  $\min \mathbf{c}(M + T)$  is achieved. Since  $\mathbf{a}_\nu$  is eligible, there exists  $R$  such that

$$\max \mathbf{c}(R) = c_\nu(R). \quad (6.6.5)$$

Also, since  $M \leq R \leq M + T$ , we may obtain  $\mathbf{c}(M + T)$  by adding at most  $T$  columns to  $\mathbf{c}(R)$ , recognizing that every time the components of the added vectors are bounded by  $a$ . Hence

$$c_j(M + T) \geq c_j(R) - aT, \quad c_j(M + T) \leq c_j(R) + aT.$$

In particular,

$$\min \mathbf{c}(M + T) = c_\nu(M + T) \geq c_\nu(R) - aT. \quad (6.6.6)$$

Suppose  $\nu_0$  is such that  $\max \mathbf{c}(M + T) = c_{\nu_0}(M + T)$ . Then by equation (6.6.5)

$$c_{\nu_0}(M + T) \leq c_{\nu_0}(R) + aT \leq c_\nu(R) + aT. \quad (6.6.7)$$

Equations (6.6.6) and (6.6.7) imply

$$\max \mathbf{c}(M + T) - \min \mathbf{c}(M + T) \leq 2aT.$$

►LEMMA 6.6.3. If all rows and columns of  $\mathbf{A}$  are eligible in  $(M, M + T)$ , then

$$\max \mathbf{c}(M + T) - \min \mathbf{r}(M + T) \leq 4aT. \quad (6.6.8)$$

*Proof.*

$$\begin{aligned} \max \mathbf{c}(M + T) - \min \mathbf{r}(M + T) &= [\max \mathbf{c}(M + T) - \min \mathbf{c}(M + T)] \\ &\quad + [\min \mathbf{c}(M + T) - \max \mathbf{r}(M + T)] \\ &\quad + [\max \mathbf{r}(M + T) - \min \mathbf{r}(M + T)] \\ &\leq 4aT + [\min \mathbf{c}(M + T) - \max \mathbf{r}(M + T)]. \end{aligned}$$

It remains to show that the last term is negative.

Let  $k_i, l_j$  denote the number of times that  $\mathbf{b}_i, \mathbf{a}_j$  are added to  $\mathbf{c}(0), \mathbf{r}(0)$  in order to form  $\mathbf{c}(M + T)$  and  $\mathbf{r}(M + T)$ , respectively.

As in Lemma 6.6.1, we see that

$$\frac{1}{M + T} \sum k_i = \frac{1}{M + T} \sum l_j = 1.$$

Let  $\mathbf{B} = \mathbf{A}'$  and thus  $b_{ij} = a_{ji}$ ; let  $v^*$  denote the value of the game with matrix  $\mathbf{B}$ . Then

$$(M + T)\mathbf{x} = \langle k_1, \dots, k_m \rangle, \quad (M + T)\mathbf{y} = \langle l_1, \dots, l_n \rangle$$

are strategies for Players I and II, respectively, of the game with matrix  $\mathbf{B}$ . Furthermore,

$$\frac{c_j(M + T)}{M + T} = \frac{c_j(0)}{M + T} + \sum_i b_{ij} \frac{k_i}{M + T} = \frac{c_j(0)}{M + T} + (\mathbf{x}\mathbf{B})_j$$

and

$$\frac{r_i(M + T)}{M + T} = \frac{r_i(0)}{M + T} + \sum_j b_{ij} \frac{l_j}{M + T} = \frac{r_i(0)}{M + T} + (\mathbf{B}\mathbf{y})_i.$$

Then

$$\begin{aligned} \frac{\min \mathbf{c}(M + T) - \max \mathbf{r}(M + T)}{M + T} &= \frac{\min \mathbf{c}(0) - \max \mathbf{r}(0)}{M + T} \\ &\quad + \min \mathbf{x}\mathbf{B} - \max \mathbf{B}\mathbf{y}. \end{aligned}$$

Also,  $\min \mathbf{c}(0) = \max \mathbf{r}(0)$  by definition of a vector system. Finally, arguing as in Lemma 6.6.1, we find that

$$\min \mathbf{x}\mathbf{B} \leq v^* \leq \max \mathbf{B}\mathbf{y},$$

or

$$\min \mathbf{x}\mathbf{B} - \max \mathbf{B}\mathbf{y} \leq 0.$$

Hence

$$\frac{\min \mathbf{c}(M + T) - \max \mathbf{r}(M + T)}{M + T} \leq 0.$$

This completes the proof of the lemma.

►LEMMA 6.6.4. For every  $\epsilon > 0$ , there exists  $T^*(\epsilon)$  such that if  $T > T^*$ , then

$$\frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} < \epsilon$$

for every vector system associated with a fixed matrix  $\mathbf{A}$ . The value of  $T^*(\epsilon)$  is independent of the vector system.

*Proof.* We shall proceed by induction. The lemma is clearly valid for  $1 \times 1$  matrices, since  $r(0) = c(0)$  implies  $r(n) \equiv c(n)$  for all  $n$ . We shall assume that the lemma is valid for all submatrices of  $\mathbf{A}$  and then show that it is true for  $\mathbf{A}$ . Since the number of submatrices is finite, we may assume that  $T^*(\epsilon/2)$  is the same for all proper submatrices of  $\mathbf{A}$  for all vector systems. In the remainder of the proof  $T^*$  will denote this number.

**SUB-LEMMA.** If in the interval  $(M, M + T^*)$  the  $v$ th row is ineligible, then

$$\max \mathbf{c}(M + T^*) - \min \mathbf{r}(M + T^*) \leq \max \mathbf{c}(M) - \min \mathbf{r}(M) + \frac{\epsilon}{2} T^*, \quad (6.6.9)$$

where  $T^* = T^*(\epsilon/2)$ . The same result holds if the  $u$ th column is ineligible.

*Proof of the Sub-lemma.* Define the new vector system:

$$\left. \begin{array}{l} \bar{\mathbf{r}}(k) = \mathbf{r}(M + k) + [\max \mathbf{c}(M) - \min \mathbf{r}(M)]\mathbf{u}, \\ \text{where } \mathbf{u} = \langle 1, 1, \dots, 1 \rangle \\ \bar{\mathbf{c}}(k) = \text{projection of } \mathbf{c}(M + k) \text{ obtained by} \\ \text{deleting the } v \text{th row} \end{array} \right\} (k \leq T^*).$$

For  $k > T^*$ , use rules (2) and (3) of the definition of a vector system (page 182) to continue the system with respect to the submatrix obtained from  $\mathbf{A}$  by deleting the  $v$ th row.

We can verify that  $\bar{\mathbf{r}}(k)$  and  $\bar{\mathbf{c}}(k)$  form a vector system. Rule (1) is satisfied as follows:

$$\begin{aligned} \min \bar{\mathbf{r}}(0) &= \min \mathbf{r}(M) + \max \mathbf{c}(M) - \min \mathbf{r}(M) \\ &= \max \mathbf{c}(M) = \max \bar{\mathbf{c}}(0), \end{aligned}$$

since the  $v$ th row is ineligible. Rules (2) and (3) are satisfied for  $k \leq T^*$  because the  $v$ th row is ineligible, and for  $k > T^*$  by construction.

Now we can prove (6.6.9). Since  $\bar{\mathbf{r}}(k)$  and  $\bar{\mathbf{c}}(k)$  are defined over a submatrix of  $\mathbf{A}$ , namely the submatrix obtained by deleting the  $v$ th row, we have by the induction hypothesis

$$\frac{\max \bar{\mathbf{c}}(T^*) - \min \bar{\mathbf{r}}(T^*)}{T^*} < \frac{\epsilon}{2}, \quad (6.6.10)$$

which is equivalent to (6.6.9).

*Proof of the Lemma.* Let  $T = (\theta + q)T^*$  ( $0 \leq \theta < 1$ ;  $q$  and  $T$  integers, with the value of  $q$  to be specified later). Let

$$I_s = [(\theta + s - 1)T^*, (\theta + s)T^*] \quad (1 \leq s \leq q).$$

*Case 1.* There exists an  $s$  such that all rows and columns of  $\mathbf{A}$  are eligible in  $I_s$ .

Let  $s'$  be the largest such  $s$ . If  $s' \leq q - 1$ , we can apply the sub-lemma to the interval  $(M, M + T^*)$  with  $M = (\theta + q - 1)T^*$  and  $M + T^* = T$ ; then

$$\begin{aligned} \max \mathbf{c}(T) - \min \mathbf{r}(T) &\leq \max \mathbf{c}[(\theta + q - 1)T^*] \\ &\quad - \min \mathbf{r}[(\theta + q - 1)T^*] + \frac{\epsilon}{2} T^*. \end{aligned} \quad (6.6.11)$$

If  $s' \leq q - 2$ , we can repeat the inequality by applying the sub-lemma to the right side of (6.6.11). Then

$$\begin{aligned} \max \mathbf{c}(T) - \min \mathbf{r}(T) &\leq \max \mathbf{c}[(\theta + q - 2)T^*] \\ &\quad - \min \mathbf{r}[(\theta + q - 2)T^*] + 2 \frac{\epsilon}{2} T^*. \end{aligned}$$

We continue iterating the inequalities until we reach  $I_{s'}$ . There the process must stop, since all rows and columns are eligible. Hence

$$\begin{aligned} \max \mathbf{c}(T) - \min \mathbf{r}(T) &\leq \max \mathbf{c}[(\theta + s')T^*] \\ &\quad - \min \mathbf{r}[(\theta + s')T^*] + (q - s') \frac{\epsilon}{2} T^* \\ &\leq \max \mathbf{c}[(\theta + s')T^*] \\ &\quad - \min \mathbf{r}[(\theta + s')T^*] + q \frac{\epsilon}{2} T^*. \end{aligned} \quad (6.6.12)$$

But in the interval  $I_s = [(\theta + s' - 1)T^*, (\theta + s')T^*]$  of length  $T^*$  all rows and columns are eligible, so that Lemma 6.6.3 may be applied to yield

$$\max \mathbf{c}[(\theta + s')T^*] - \min \mathbf{r}[(\theta + s')T^*] \leq 4aT^*. \quad (6.6.13)$$

On substituting this result in (6.6.12), we obtain

$$\max \mathbf{c}(T) - \min \mathbf{r}(T) \leq 4aT^* + q \frac{\epsilon}{2} T^*.$$

*Case 2.* There exists no  $s$  such that all rows and columns of  $\mathbf{A}$  are eligible in  $I_s$ . Then the sub-lemma may be applied as in (6.6.11) and the result iterated until  $s = 1$ . We obtain

$$\begin{aligned}\max \mathbf{c}(T) - \min \mathbf{r}(T) &\leq \max \mathbf{c}(\theta T^*) - \min \mathbf{r}(\theta T^*) + q \frac{\epsilon}{2} T^* \\ &\leq 2\theta a T^* + q \frac{\epsilon}{2} T^* + M,\end{aligned}$$

where  $M = \max |\mathbf{c}(0)| + \max |\mathbf{r}(0)|$ .

The last inequality holds because  $\mathbf{c}(\theta T^*)$  is obtained by adding  $\theta T^*$  columns to  $\mathbf{c}(0)$ , and we know that  $|a_{ij}| \leq a$ . Hence

$$|\max \mathbf{c}(\theta T^*)| \leq a\theta T^* + \max |\mathbf{c}(0)|.$$

Similarly,

$$|\min \mathbf{r}(\theta T^*)| \leq a\theta T^* + \max |\mathbf{r}(0)|.$$

Finally, since  $\theta < 1$ , if we choose  $T^* > M/2a$ , we obtain

$$\max \mathbf{c}(T) - \min \mathbf{r}(T) \leq 4aT^* + q \frac{\epsilon}{2} T^*. \quad (6.6.14)$$

Clearly, equation (6.6.14) is valid for both cases. Then

$$\frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} \leq \frac{4a + q(\epsilon/2)}{q} \leq \frac{4a}{q} + \frac{\epsilon}{2}.$$

If we choose  $q$  so large that  $4a/q < \epsilon/2$ , then for all  $T > qT^*$

$$\frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} < \epsilon.$$

*Proof of Theorem 6.6.1.* By Lemma 6.6.4

$$\overline{\lim} \left[ \frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} \right] \leq 0,$$

and by Lemma 6.6.1

$$\underline{\lim} \left[ \frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} \right] \geq 0.$$

Therefore

$$\lim \left[ \frac{\max \mathbf{c}(T) - \min \mathbf{r}(T)}{T} \right] = 0. \quad (6.6.15)$$

But

$$\overline{\lim} \left[ \min \frac{\mathbf{r}(T)}{T} \right] \leq v \leq \underline{\lim} \left[ \max \frac{\mathbf{c}(T)}{T} \right],$$

showing that each limit in (6.6.15) exists separately; hence

$$\lim_T \left[ \min \frac{\mathbf{r}(T)}{T} \right] = \lim_T \left[ \max \frac{\mathbf{c}(T)}{T} \right] = v.$$

**\*6.7 A differential-equations method for determining the value of a game.** The method discussed in Sections 6.5 and 6.6 was discrete in concept and in operation. In contrast, our next approximation method involves continuous alteration of the strategies to take advantage of the opponent's errors. The strategies vary continuously in time and satisfy certain differential equations which express the way each player adjusts his play in reaction to his opponent's play. When solved, these differential equations furnish a one-parameter family  $\mathbf{y}(t)$  of strategies for each player, with the property that as time goes to infinity all limit points of  $\mathbf{y}(t)$  become optimal.

Unlike the discrete method, this continuous method enables us to establish precise bounds on the rates of convergence. This advantage, however, and the method itself are more of theoretical interest than of practical value, since in most cases the differential equations are difficult to solve.

Since any matrix game may be imbedded in an equivalent symmetric game (see Theorem 2.6.1), we shall consider here the symmetric game only.

Let  $\Gamma$  be a symmetric game with skew-symmetric  $n \times n$  matrix  $\mathbf{A}$ . We consider a one-parameter family of strategies  $\mathbf{y}(t)$  as functions of the real variable  $t$ . Let us form the following functions of  $\mathbf{y}(t)$ :

$$\begin{aligned} u_i[\mathbf{y}(t)] &= \sum_{j=1}^n a_{ij}y_j(t), & \phi(u_i) &= \max (0, u_i), \\ \phi(\mathbf{y}) &= \sum_{i=1}^n \phi(u_i), & \psi(\mathbf{y}) &= \sum_{i=1}^n \phi^2(u_i). \end{aligned} \tag{6.7.1}$$

Let  $\mathbf{y}^0$  be any fixed strategy for  $\Gamma$ . We now consider the following system of differential equations.

$$\left. \begin{aligned} \frac{dy_j(t)}{dt} &= \phi[u_j(t)] - \phi[\mathbf{y}(t)]y_j(t), \\ y_j(0) &= y_j^0, \end{aligned} \right\} \quad (j = 1, \dots, n). \tag{6.7.2}$$

The reason we choose these particular differential equations is as follows. Suppose that  $\phi(u_j) > 0$ ; then by playing the  $j$ th row Player I obtains more than the value of the game, provided Player II plays  $\{y_j\}$  (remember  $v = 0$ ). If Player II now changes his strategy by increasing  $y_j$  to 1, the return to Player I becomes  $a_{jj}$ , which is zero. Hence it is to Player II's advantage to increase  $y_j$  whenever  $\phi(u_j) > 0$ . This situation is accounted for by the first term of the differential equation, which compels  $y_j$  to increase whenever  $\phi(u_j) > 0$ . The second term serves the role of a normalizing factor; it maintains the relations  $\sum y_j(t) \equiv 1$ , provided  $\mathbf{y}^0$  was a strategy to start with.

Since the right side is continuous, it is known from the theory of differential equations that such a system has at least one solution. We shall prove that if  $y_j(t)$  is any solution, and  $\{t_i\}$  is a sequence converging to infinity, then any limit point of  $\mathbf{y}(t_i)$  is an optimal strategy for Player II.

►LEMMA 6.7.1. Let  $y_j(t)$  ( $j = 1, \dots, n$ ) be a solution to the system of equations (6.7.2). Then  $\langle y_j(t) \rangle$  is a strategy for any fixed  $t$ .

*Proof.* We prove first that  $y_j(t) \geq 0$ . Assume that this is false. Then there exists  $t_1$  such that  $y_j(t_1) < 0$ . Let  $t_0$  be the largest value of  $t$  less than  $t_1$  for which  $y_j(t_0) \geq 0$ . Actually,  $y_j(t_0) = 0$ . Then if  $t_0 < t \leq t_1$ , it follows that  $y_j(t) < 0$ ; but  $\phi(u_j) \geq 0$  and  $\phi(\mathbf{y}) \geq 0$  always. Hence  $\phi(u_j) - \phi(\mathbf{y}) \geq 0$ , and thus  $dy_j/dt \geq 0$  by (6.7.2). Applying the mean value theorem, we obtain

$$y_j(t_1) = y_j(t_0) + y'_j(\tau)(t_1 - t_0) \geq y_j(t_0) = 0,$$

which contradicts the assumption  $y_j(t_1) < 0$ .

We prove next that

$$\sum_{j=1}^n y_j(t) = 1.$$

Observe that

$$\begin{aligned} \frac{d}{dt} \left( 1 - \sum_{j=1}^n y_j(t) \right) &= - \sum_{j=1}^n \frac{dy_j}{dt} = - \sum_{j=1}^n \phi(u_j) + \phi(\mathbf{y}) \sum_{j=1}^n y_j \\ &= -\phi(\mathbf{y}) \left( 1 - \sum_{j=1}^n y_j(t) \right). \end{aligned} \quad (6.7.3)$$

Integrating equation (6.7.3) from 0 to  $t$ , taking account of the initial condition, we obtain

$$1 - \sum_{j=1}^n y_j(t) = - \int_0^t \phi(\mathbf{y}) \left( 1 - \sum_{j=1}^n y_j(\tau) \right) d\tau.$$

The equation has an explicit integral

$$1 - \sum_{j=1}^n y_j(t) = c \exp \left[ - \int_0^t \phi(y) dy \right],$$

and the initial conditions imply  $c = 0$ .

►LEMMA 6.7.2.

$$[\psi(\mathbf{y})]^{1/2} \leq \phi(\mathbf{y}) \leq n^{1/2}[\psi(\mathbf{y})]^{1/2}.$$

*Proof.* The first inequality is immediate, since

$$[\phi(\mathbf{y})]^2 = [\sum \phi(u_j)]^2 \geq \sum [\phi(u_j)]^2 = \psi(\mathbf{y}).$$

By Schwarz's inequality,

$$\phi(\mathbf{y}) = \sum_{j=1}^n 1 \cdot \phi(u_j) \leq \left[ \sum_{j=1}^n 1 \right]^{1/2} \left[ \sum_{j=1}^n [\phi(u_j)]^2 \right]^{1/2} = n^{1/2} [\psi(\mathbf{y})]^{1/2},$$

as desired.

► **THEOREM 6.7.1.**

$$\psi[\mathbf{y}(t)] \leq \frac{C_0}{(C_2 + t)^2}; \quad \phi[\mathbf{y}(t)] \leq \frac{C_1}{C_2 + t}.$$

*Proof.* If  $\phi(u_i) > 0$  for a given  $t$ , then

$$\begin{aligned} \frac{d\phi(u_i)}{dt} &= \sum_{j=1}^n a_{ij} \frac{dy_j}{dt} = \sum_{j=1}^n a_{ij} [\phi(u_j) - \phi(\mathbf{y})y_j] \\ &= \sum_{j=1}^n a_{ij} \phi(u_j) - \phi(\mathbf{y}) \phi(u_i). \end{aligned}$$

If we multiply both sides by  $\phi(u_i)$  and sum, we obtain the following equation, valid even if  $\phi(u_i) = 0$ :

$$\begin{aligned} \sum_{i=1}^n \phi(u_i) \frac{d\phi(u_i)}{dt} &= \sum_{i,j} a_{ij} \phi(u_i) \phi(u_j) - \phi(\mathbf{y}) \sum_{i=1}^n \phi^2(u_i) \\ &= -\phi(\mathbf{y}) \psi(\mathbf{y}). \end{aligned} \tag{6.7.4}$$

The sum over  $i, j$  vanishes because  $\mathbf{A}$  is skew-symmetric; that is,  $a_{ij} = -a_{ji}$ . Equation (6.7.4) is equivalent to

$$\frac{1}{2} \frac{d\psi[\mathbf{y}(t)]}{dt} = -\phi[\mathbf{y}(t)] \psi[\mathbf{y}(t)]. \tag{6.7.5}$$

Therefore  $\psi[\mathbf{y}(t)]$  is a decreasing function of  $t$ . By Lemma 6.7.2, we know also that it vanishes only when  $\phi[\mathbf{y}(t)]$  vanishes, and further that

$$\frac{1}{2} \frac{d\psi}{dt} \leq -\psi^{1/2} \psi.$$

Then

$$\frac{1}{2} \frac{d\psi}{\psi^{3/2}} \leq -dt, \quad -\psi^{-1/2}[\mathbf{y}(t)] + \psi^{-1/2}[\mathbf{y}(0)] \leq -t.$$

If  $C_2 = \psi^{-1/2}[\mathbf{y}(0)]$ , we have  $\psi^{-1/2}[\mathbf{y}(t)] \geq C_2 + t$  or

$$\psi^{1/2}[\mathbf{y}(t)] \leq \frac{1}{C_2 + t}. \quad (6.7.6)$$

The last equation implies that for every  $i$ ,

$$\phi(u_i) \leq \frac{n^{1/2}}{C_2 + t}.$$

Since  $\phi(u_i) = \max(0, u_i)$ , we see that

$$u_i = \sum a_{ij} y_j(t) \leq \frac{n^{1/2}}{C_2 + t}.$$

Hence, if a sequence  $\{t_m\} \rightarrow \infty$ , then  $\sum a_{ij} y_j(t_m)$  has a limit superior bounded above by zero, which is the value of the game. Thus any limit point of the sequence  $\mathbf{y}(t_m)$  is an optimal strategy for Player II, and also for Player I since the game is symmetric. We note that the yields

$$\sum a_{ij} y_j(t_m)$$

reduce to less than or equal to  $v = 0$  at least as fast as  $n^{1/2}/(C_2 + t)$ .

### 6.8 Problems.

1. Use the simplex method to maximize

$$2\xi_1 + 4\xi_2 + \xi_3 + \xi_4$$

subject to the conditions

$$\xi_1 + 3\xi_2 + \xi_4 \leq 4,$$

$$2\xi_1 + \xi_2 \leq 3,$$

$$\xi_2 + 4\xi_3 + \xi_4 \leq 3.$$

2. Three processes are available for producing two final products, corn and hogs. Process  $A$  has hogs as a final product but uses corn as an intermediate product (for feeding the hogs); Process  $B$  is concerned only with growing corn; and Process  $C$  has corn and hogs as joint products. Corn sells for \$20 per unit (ton), and hogs also sell for \$20 per unit (head). Process  $A$  uses 100 unit-months of labor and 10 acres of land per unit output; Process  $B$  uses 25 unit-months of labor and 50 acres of land per unit output; and Process  $C$  uses 50 unit-months of labor and 40 acres of land

per unit output. There are 50 unit-months of labor available and 52.5 acres of land. Assuming constant returns to scale, find the levels at which Processes  $A$ ,  $B$ , and  $C$  must be employed in order to achieve maximum profits.

3. Use the simplex method to solve the transportation problem for the following array:

<i>Sources</i>	<i>Destinations</i>	<i>Unit shipping costs</i>					
$a_1 = 3$	$b_1 = 3$	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$
$a_2 = 4$	$b_2 = 3$	$a_1$	5	3	7	3	8
$a_3 = 2$	$b_3 = 6$	$a_2$	5	6	12	5	7
$a_4 = 8$	$b_4 = 2$	$a_3$	2	8	3	4	8
	$b_5 = 1$	$a_4$	9	6	10	5	10
	$b_6 = 2$						9

4. If a system of  $m$  linear equations in  $n$  nonnegative variables has a solution, then a solution exists in which  $k$  variables are positive and  $n - k$  are zero, with  $k \leq \min(m, n)$ .

5. In the warehouse model of Section 5.7, solve explicitly the special case in which  $b_i > c_i$ ,  $b_{i+1} > c_i$ , and  $c_i \geq 0$  for all  $i$ .

6. Consider a transportation problem with availabilities at the sources given by  $c_i > 0$  ( $i = 1, \dots, n$ ) and demands at the destinations given by  $b_j > 0$  ( $j = 1, \dots, m$ ), such that

$$\sum_{i=1}^n c_i = \sum_{j=1}^m b_j.$$

Show that the rank of the constraint matrix  $\mathbf{A}$  of page 137 is exactly  $m + n - 1$ .

7. Solve the diet problem

$$\min(\mathbf{c}, \mathbf{x}) \text{ subject to } \mathbf{Ax} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0},$$

where

$$\mathbf{A} = \begin{vmatrix} 2 & 1 & 7 & 1 & 0 & 4 \\ 3 & 0 & 2 & 8 & 6 & 4 \\ 1 & 4 & 7 & 3 & 5 & 0 \end{vmatrix}, \quad \mathbf{b} = \begin{vmatrix} 1 \\ 1 \\ 1 \end{vmatrix},$$

and

$$\mathbf{c} = \langle 2 \ 3 \ 4 \ 1 \ 2 \ 3 \rangle.$$

Use the method of artificial variables and then the dual simplex method.

8. Minimize

$$z_1 + z_2 + z_3 + \lambda x,$$

where

$$z_1 + z_2 + z_3 + x \geq 10$$

$$z_1 + z_2 + x \geq 7 \quad 0 \leq z_i \leq 5$$

$$z_2 + z_3 + x \geq 9 \quad x \geq 0,$$

and  $\lambda$  is a positive parameter.

9. Prove that degeneracy in the simplex computation for the transportation problem can occur only if a partial sum of the  $c_i$  equals a partial sum of the  $b_j$ .

10. Set up the caterer's problem as a transportation problem.

11. A factory has 100 items on hand which may be shipped to an outlet at the cost of \$1 apiece to meet an uncertain demand  $d$  which is uniformly distributed between 70 and 80. In the event that the demand exceeds the supply, it is necessary to meet the unsatisfied demand by purchases at the local market at \$2 apiece. Set up the equations and find the shipping policy that minimizes expected costs.

12. Use the labeling algorithm to find a maximal flow and a minimal cut for the network diagrammed below. Reverse the arc orientations, interchange source and sink, and locate another minimal cut by labeling.

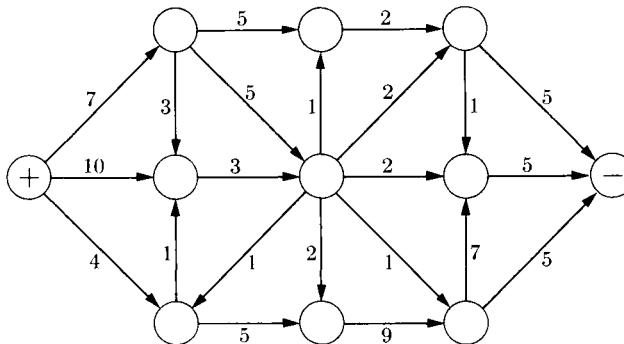


FIGURE 6.6

13. Consider an optimal assignment problem in which there are only two ratings: i.e., man  $i$  in job  $j$  is either good or bad. Show that such a problem is a maximal network flow problem.

14. Show how a flow problem with capacity constraints on nodes can be converted to one with capacity constraints on arcs only.

15. For a given unoriented network, the capacity constraint on an arc  $p_i p_j$  is of the form  $x_{ij} + x_{ji} \leq c_{ij}$ . Reduce this to the case of an oriented network with constraints  $x_{ij} \leq c_{ij}$ . How is the statement of the max-flow, min-cut theorem for the original network affected?

16. The matrix  $\mathbf{A}$  of the constraints (6.4.1) is the incidence matrix of nodes *vs.* arcs of the network. Assuming that all arcs are present, prove that the rank of  $\mathbf{A}$  is  $n - 1$ .

17. Write out the dual of the maximal flow problem as described in (6.4.1) and (6.4.2), and show that the labeling process yields an integral solution to the dual problem.

18. Suppose that  $\{S_i, \bar{S}_i\}$ , where  $i = 1, 2$ , represent two minimal cuts in partition form; i.e., the set of arcs leading from  $S_i$  to its complement  $\bar{S}_i$  is a minimal cut. Show that  $\{S_1 \cup S_2, \bar{S}_1 \cap \bar{S}_2\}$  and  $\{S_1 \cap S_2, \bar{S}_1 \cup \bar{S}_2\}$  also represent minimal cuts.

19. Let  $\mathbf{P}_j$  and  $\mathbf{Q}$  denote  $n + 1$  vectors in  $E^m$  ( $m < n$ ) satisfying the nondegeneracy Assumption 6.1.1. Let  $X$  describe the set of all  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  satisfying

$$\sum_{i=1}^n x_i \mathbf{P}_i = \mathbf{Q}$$

and  $\mathbf{x} \geq \mathbf{0}$ . Define

$$\phi_1 = \sum_{i=1}^n \alpha_i x_i, \quad \phi(\mathbf{x}) = \phi_1(\mathbf{x}) + \beta \sum_{i=1}^n \delta(x_i)$$

for  $\mathbf{x}$  in  $X$ , where  $\beta \geq 0$  and  $\delta(x_i)$  equals 1 or 0 according as  $x_i > 0$  or  $x_i = 0$ . Prove that any extreme point of  $X$  which minimizes  $\phi_1(\mathbf{x})$  also minimizes  $\phi(\mathbf{x})$ , and conversely that any minimum point of  $\phi(\mathbf{x})$  is also a minimum point of  $\phi_1(\mathbf{x})$ .

20. Suppose that the nondegeneracy assumption of Problem 19 is removed, while the other hypotheses remain in effect. Show that if

$$\inf_{\mathbf{x} \in X} \sum_{j=1}^n [\alpha_j x_j + \beta_j \delta(x_j)] = \mu > -\infty \quad (\beta_j \geq 0),$$

then there exists an extreme point  $\mathbf{x}^*$  of  $X$  such that

$$\sum_{j=1}^n \alpha_j x_j^* + \beta_j \delta(x_j^*) = \mu.$$

21. Minimize

$$\sum_{i=1}^N u_i$$

subject to restrictions

$$u_i + u_{i+1} \geq \alpha_i \quad (i = 1, \dots, N - 1),$$

$$u_N + u_1 \geq \alpha_N.$$

22. Use the labeling algorithm and Problem 16 to prove the max-flow, min-cut theorem (Theorem 5.10.1).
23. Let  $a_1$  and  $a_2$  be two arcs of a network flow with capacities  $c_1$  and  $c_2$ , respectively, and let  $\Delta_1$  be the change in maximal flow value effected by increasing  $c_1$  one unit while holding other capacities fixed. Similarly define  $\Delta_2$ . Let  $\Delta_{12}$  be the change in flow value effected by increasing  $c_1$  and  $c_2$  one unit each. Prove that
- if  $a_1$  and  $a_2$  both lead from the source, then  $\Delta_1 + \Delta_2 \geq \Delta_{12}$ ;
  - if  $a_1$  leads from the source and  $a_2$  leads into the sink, then  $\Delta_1 + \Delta_2 \leq \Delta_{12}$ .

#### NOTES AND REFERENCES TO CHAPTER 6

**§ 6.1.** The practical importance of linear programming theory depends to a large extent on the simplex algorithm described in this section. Dantzig attributes the method to the famous fertile imagination of von Neumann, who reportedly proposed it in a casual conversation between them. It was Dantzig, however, who exploited the suggestion and developed the method, which is now recognized as the forte of linear programming methodology.

A thorough elementary summary and illustration of the simplex algorithm and some of its modifications is contained in Charnes, Cooper, and Henderson [46]. A discussion of the method and further illustrations are included in Churchman, Ackoff, and Arnoff [50]. Some other expository articles on the algorithm are given by Eisemann [74] and Dantzig [56]. Comprehensive books on the subject are being written by Dantzig and by Charnes and Cooper; see also Gass [103]. Dorfman, Samuelson, and Solow [65] offers an exhaustive introductory discussion of the concepts and methodology of linear programming and related economic analysis.

The use of slack variables was initially proposed by Dantzig, and the device of artificial variables was also introduced by Dantzig. The problem of handling degeneracies in the simplex algorithm was solved by Charnes [44] and independently by Dantzig, Orden, and Wolfe [59]; our discussion follows Charnes.

**§ 6.2.** The dual simplex algorithm presented here is that of Lemke [165]. Wagner [243] gives an illuminating exposition of the algorithm and discusses the general usefulness of the dual theorem as a computational aid. There exist techniques which combine the primal and dual simplex algorithms. For a discussion of these techniques, which are useful for some problems, we direct the reader to Lemke [165] and Dantzig [57]. Special adjustments of the primal and dual algorithm to deal with secondary constraints and parametric programming problems are discussed by Gass [103], Wagner [243], and Dantzig [57].

**§ 6.3.** The illustration of this section was offered by Wagner.

**§ 6.4.** The elegant algorithm of this section, valid for network flow problems and numerous varieties of the transportation problem, is due to Ford and Fulkerson [85]. We have closely followed their presentation. For other transportation algorithms, the reader is referred to [90], [91], and [160].

**§ 6.5.** The iterative convergence procedure for determining the value of a game was proposed by Brown [41]. It was Robinson [208] who rigorously established the convergence. Variations of the convergence procedure have been given by von Neumann [241] and Brown [40]. Continuous analogs of the scheme applying to infinite games are discussed by Danskin [51].

**§ 6.6.** The proof of the convergence of the Brown method appears to be very delicate. The rate of convergence needs further study, along with such questions as whether the convergence is affected if the past strategies are assigned unequal weightings.

**§ 6.7.** The differential-equations method, which is due to Brown and von Neumann [42], is introduced primarily to shed light on the discrete procedure of § 6.5. Modifications of this method have been suggested by Bellman (unpublished).

**§ 6.8.** Problem 2 is due to Allen [2]. Problems 14–18 were suggested by Ford and Fulkerson [85]. Problems 19 and 20 are due to Hirsch and Dantzig [123]. Problem 21 is due to Gross [115]. The idea of Problem 23 was noted by Shapley.

## CHAPTER 7

### NONLINEAR PROGRAMMING

Nonlinear programming is an extension of the theory of linear programming to deal with problems of maximizing nonlinear functions subject to general constraints. Unfortunately, it seems difficult to obtain general qualitative characterizations of the solutions to nonlinear programming problems. Some success, however, has been achieved where the function to be maximized is concave (or the function to be minimized is convex) and the constraints limit the free variables to ranging over a convex set. Fortunately, many economic and management problems arising in practice involve functions which are naturally convex or concave. Heuristically, the convexity assumption refers to a situation of increasing returns per unit increase of input, and the concavity assumption to a situation of decreasing returns per unit increase of input. Furthermore, it is common in theoretical economic analysis to suppose that the utility functions representing individual desire are concave. (See Section 8.5.)

From a mathematical point of view, the principal reason that convex or concave functions yield to analysis hinges on the fact that such functions can be approached by linear functions with certain inequalities preserved; in geometric terms, there are supporting planes to the graph of the function which keep the whole graph of the function on one side of the hyperplane.

The primary object of this chapter is to demonstrate the analogs of the duality theorem and the equivalence of game problems to linear programming problems for the case of nonlinear programming. This is accomplished in Sections 7.1 and 7.8.

An elaborate interindustry nonlinear activity analysis model is described in Section 7.2. The model is of interest because of its interpretations and because of the special mathematical methods used in its analysis.

An approximate solution to the concave programming problem can be obtained by solving an associated system of gradient differential equations. The solutions to these equations may be regarded as expressing the perturbation mechanism of changing feasible vectors into feasible vectors with larger return. These ideas are discussed in Section 7.3.

The important theory of conjugate convex functions initiated by Fenchel is developed in Sections 7.5 through 7.7. This elegant and interesting mathematical theory is applied in the concluding sections to the duality theory of nonlinear programming and to the structure of convex sets.

**7.1 Concave programming.** Nonlinear programming deals with problems such as: Maximize a function  $g(\mathbf{x})$  of  $n$  real variables constrained by  $m$  inequalities  $f_j(\mathbf{x}) \geq 0$  ( $j = 1, \dots, m$ ) and  $\mathbf{x} \in X$ , where  $X$  is a convex subset of  $E^n$ . In most applications,  $X$  is identified with the nonnegative orthant of  $E^n$ . The problem is called concave programming if  $g(\mathbf{x})$  and  $f_j(\mathbf{x})$  are all concave functions of  $\mathbf{x}$ . Unless there is a statement to the contrary, we assume from here on that the functions  $g(\mathbf{x})$  and  $f_j(\mathbf{x})$  are concave in the region  $\mathbf{x} \in X$ .

Kuhn and Tucker were the first to give a formal discussion of various mathematical aspects of concave programming problems. They distinguish a class of “proper” solutions, which are essentially solutions that can be recognized unambiguously by differential conditions; and their proofs, depending crucially on differential methods, apply only to this class of proper solutions. The methods and proofs given below do not involve such restrictions. Our results, however, are complementary to theirs rather than an extension of theirs; hence we shall also discuss the differential approach later in this section.

Under suitable assumptions the problem of characterizing the maximum points  $\mathbf{x}^0$  of the concave programming problem described above can be transformed into an equivalent game or saddle value problem.

We shall prove first (Theorem 7.1.1 below) that a particular vector  $\mathbf{x}^0$  maximizes  $g(\mathbf{x})$  subject to the constraints  $\{f_j(\mathbf{x})\} = \mathbf{F}(\mathbf{x}) \geq 0$  (the vector function as indicated) and  $\mathbf{x} \in X$ , if and only if there exists some  $\mathbf{u}^0$  with nonnegative components which, for the function

$$\phi(\mathbf{x}, \mathbf{u}) = g(\mathbf{x}) + (\mathbf{u}, \mathbf{F}(\mathbf{x})) \quad (7.1.1)$$

satisfies

$$\phi(\mathbf{x}, \mathbf{u}^0) \leq \phi(\mathbf{x}^0, \mathbf{u}^0) \leq \phi(\mathbf{x}^0, \mathbf{u}) \quad (7.1.2)$$

for all  $\mathbf{u} \geq 0$  and  $\mathbf{x} \in X$ . The pair  $\{\mathbf{x}^0, \mathbf{u}^0\}$  will be called a saddle point of the kernel  $\phi(\mathbf{x}, \mathbf{u})$ . A feasible vector  $\mathbf{x}^0$ , i.e., a point at which  $\mathbf{x}$  satisfies the constraints  $\mathbf{x} \in X$  and  $\mathbf{F}(\mathbf{x}) \geq 0$ , is called optimal if  $\mathbf{x}^0$  maximizes  $g(\mathbf{x})$  among all feasible vectors.

The economic interpretation of the maximization as suggested by Kuhn and Tucker can best be stated in relation to the game with pay-off kernel  $\phi(\mathbf{x}, \mathbf{u})$ . The constraints are usually of the form  $f_j(\mathbf{x}) = \xi_j - k_j(\mathbf{x})$ , with  $\xi$  a constant vector and  $k_j$  a convex function of its argument in the region  $\mathbf{x} \geq 0$ . The vector  $\mathbf{x}$  represents the activity level of the various operations. Interpret  $\xi$  as the available (vector) amount of primary resources and  $\mathbf{k}(\mathbf{x})$  as the (vector) amount actually used, so that  $\mathbf{F}(\mathbf{x})$  is the unused balance of primary resources. The function  $g(\mathbf{x})$  is a measure of the resulting output of the desired commodity. The Lagrange multiplier  $\mathbf{u}$  appearing in the definition of  $\phi(\mathbf{x}, \mathbf{u})$  may be thought of as a price vector of the primary resources relative to a unit price for the desired commodity. The

kernel  $\phi(\mathbf{x}, \mathbf{u})$  can be interpreted as the combined output of the desired commodity and the unused balance of the primary resources. The existence of a saddle point expresses in a sense an equilibrium between prices of the available resources and the value of the output such that the use of the primary resources is controlled by the manager, while prices are generated or promulgated by some independent agency (see Chapter 8).

The first fundamental theorem may be stated as follows.

► **THEOREM 7.1.1.** Let  $g(\mathbf{x})$  and  $\{f_i(\mathbf{x})\} = \mathbf{F}(\mathbf{x})$  all be concave functions defined on the convex set  $X$  of  $E^n$ , with the property that if  $\mathbf{u}$  is a non-negative vector not identically zero, then there exists some  $\mathbf{x} \in X$  such that  $(\mathbf{u}, \mathbf{F}(\mathbf{x})) > 0$ . If  $\mathbf{x}^0$  is a point at which  $g(\mathbf{x})$  achieves its maximum subject to the constraints  $\mathbf{x} \in X$  and  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ , that is,  $\mathbf{x}^0$  is optimal, then there exists a choice of  $\mathbf{u}^0 \geq \mathbf{0}$  such that for  $\phi(\mathbf{x}, \mathbf{u}) = g(\mathbf{x}) + (\mathbf{u}, \mathbf{F}(\mathbf{x}))$ ,

$$\phi(\mathbf{x}, \mathbf{u}^0) \leq \phi(\mathbf{x}^0, \mathbf{u}^0) \leq \phi(\mathbf{x}^0, \mathbf{u}) \quad (7.1.2)$$

for all  $\mathbf{x} \in X$ ,  $\mathbf{u} \geq \mathbf{0}$ , and  $(\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) = 0$ . Conversely, if  $\{\mathbf{x}^0, \mathbf{u}^0\}$  satisfies (7.1.2), then  $\mathbf{x}^0$  maximizes  $g(\mathbf{x})$  among all  $\mathbf{x}$  satisfying  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$  and  $\mathbf{x} \in X$ .

*Remark 7.1.1.* The hypothesis of the theorem is satisfied if there exists a vector  $\mathbf{x} \in X$  such that  $\mathbf{F}(\mathbf{x})$  has all strictly positive components. (In this connection, see Problem 12.) In general, however, any  $\mathbf{x} \in X$  satisfying the hypothesis for a given  $\mathbf{u}$  need not satisfy the constraints  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ .

*Proof.* First we shall show that if  $\{\mathbf{x}^0, \mathbf{u}^0\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ , then  $\mathbf{x}^0$  maximizes  $g(\mathbf{x})$  subject to the constraints  $\mathbf{x} \in X$ ,  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ .

Substituting (7.1.1) in (7.1.2), we have

$$g(\mathbf{x}) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x})) \leq g(\mathbf{x}^0) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) \leq g(\mathbf{x}^0) + (\mathbf{u}, \mathbf{F}(\mathbf{x}^0)) \quad (7.1.3)$$

for all  $\mathbf{x} \in X$ ,  $\mathbf{u} \geq \mathbf{0}$ .

Since the right-hand inequality holds for any  $\mathbf{u} \geq \mathbf{0}$ , it follows that

$$\mathbf{F}(\mathbf{x}^0) \geq \mathbf{0}, \quad (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) = 0. \quad (7.1.4)$$

Then the left-hand inequality of (7.1.3) becomes

$$g(\mathbf{x}) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x})) \leq g(\mathbf{x}^0) \quad (\mathbf{x} \in X). \quad (7.1.5)$$

Therefore, for any  $\mathbf{x} \in X$  such that  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ , we have

$$g(\mathbf{x}) \leq g(\mathbf{x}) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x})) \leq g(\mathbf{x}^0),$$

which shows  $\mathbf{x}^0$  to be optimal.

Next we shall prove that, subject to the hypothesis of the theorem, if  $\mathbf{x}^0$  is optimal, then there exists a value  $\mathbf{u}^0$  such that  $\{\mathbf{x}^0, \mathbf{u}^0\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ .

Consider two sets  $A$  and  $B$ , in  $(m + 1)$ -dimensional space, defined by

$$A = \left\{ \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \mid \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \leq \begin{pmatrix} g(\mathbf{x}) \\ \mathbf{F}(\mathbf{x}) \end{pmatrix} \quad \text{for some } \mathbf{x} \in X \right\},$$

$$B = \left\{ \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \mid \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \gg \begin{pmatrix} g(\mathbf{x}^0) \\ \mathbf{0} \end{pmatrix} \right\},$$

where  $y^0$  is obviously a scalar and  $\mathbf{y}$  is a vector of  $m$  components.

Since  $g(\mathbf{x})$  and  $\mathbf{F}(\mathbf{x})$  are concave, the set  $A$  is convex. Also  $B$  is clearly convex, consisting of the interior of an orthant with vertex at

$$\begin{pmatrix} g(\mathbf{x}^0) \\ \mathbf{0} \end{pmatrix}$$

By the optimality of  $\mathbf{x}^0$ , sets  $A$  and  $B$  have no vector in common. Hence, by the separation theorem of convex sets (Theorem B.1.2), there is a vector

$$\begin{pmatrix} v_0 \\ \mathbf{v} \end{pmatrix} \neq 0$$

such that

$$v_0 y_0 + (\mathbf{v}, \mathbf{y}) \leq v_0 z_0 + (\mathbf{v}, \mathbf{z}) \quad \text{for all } \begin{pmatrix} y_0 \\ \mathbf{y} \end{pmatrix} \in A, \quad \begin{pmatrix} z_0 \\ \mathbf{z} \end{pmatrix} \in B. \quad (7.1.6)$$

By the definition of  $B$ , (7.1.6) implies

$$\begin{pmatrix} v_0 \\ \mathbf{v} \end{pmatrix} > \mathbf{0}. \quad (7.1.7)$$

Since

$$\begin{pmatrix} g(\mathbf{x}^0) \\ \mathbf{0} \end{pmatrix}$$

is on the boundary of  $B$ , we have, by the definition of  $A$ ,

$$v_0 g(\mathbf{x}) + (\mathbf{v}, \mathbf{F}(\mathbf{x})) \leq v_0 g(\mathbf{x}^0) \quad (\mathbf{x} \in X). \quad (7.1.8)$$

It follows that  $v_0 > 0$ . Otherwise, by (7.1.7) and (7.1.8)

$$\mathbf{v} > \mathbf{0} \quad \text{and} \quad (\mathbf{v}, \mathbf{F}(\mathbf{x})) \leq 0 \quad (\mathbf{x} \in X),$$

which contradicts the hypothesis of the theorem.

Let  $\mathbf{u}^0 = \mathbf{v}/v_0$ . Then

$$\mathbf{u}^0 \geq \mathbf{0} \quad (7.1.9)$$

and

$$g(\mathbf{x}) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x})) \leq g(\mathbf{x}^0) \quad (\mathbf{x} \in X). \quad (7.1.10)$$

Putting  $\mathbf{x} = \mathbf{x}^0$  in (7.1.10), we get

$$(\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) \leq 0.$$

On the other hand, we have

$$\mathbf{F}(\mathbf{x}^0) \geq \mathbf{0}. \quad (7.1.11)$$

Hence

$$(\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) = 0. \quad (7.1.12)$$

Equations (7.1.9–12) in conjunction show that  $\{\mathbf{x}^0, \mathbf{u}^0\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ .

The hypothesis of Theorem 7.1.1, weak as it is, may be further slightly weakened by introducing unnatural conditions. However, some sort of nondegeneracy condition must be expressed in the hypothesis or else the theorem is false. Consider the example

$$g(\mathbf{x}) = x,$$

$$\mathbf{F}(\mathbf{x}) = -x^2.$$

Clearly,  $x = 0$  will be the optimal solution to the problem. But  $\phi(\mathbf{x}, \mathbf{u}) = x - ux^2$  has no saddle point.

In the case in which the constraints are defined by linear inequalities, that is,  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$  is identical with  $\mathbf{Ax} \leq \mathbf{b}$  for some matrix  $\mathbf{A}$  and vector  $\mathbf{b}$ , it is possible to establish the conclusion of Theorem 7.1.1 without any extra hypothesis regarding the constraint set.

► **THEOREM 7.1.2.** Let  $g(\mathbf{x})$  represent a concave function defined for all  $\mathbf{x}$ . If  $\mathbf{x}^0$  is a point at which  $g(\mathbf{x})$  achieves its maximum with respect to all  $\mathbf{x}$  satisfying the restriction

$$\mathbf{x} \geq \mathbf{0}, \mathbf{Ax} \leq \mathbf{b},$$

then there exists a value  $\mathbf{u}^0$  ( $\mathbf{u}^0 \geq \mathbf{0}$ ) such that  $\{\mathbf{x}^0, \mathbf{u}^0\}$  is a saddle point of the function

$$\phi(\mathbf{x}, \mathbf{u}) = g(\mathbf{x}) + (\mathbf{u}^0, \mathbf{b} - \mathbf{Ax}),$$

and conversely.

The proof is an adaptation of the arguments of Theorem 5.3.1, utilizing the natural extension of Lemma 5.3.1 to this case. The details are omitted.

If sufficient differentiability is assumed for the functions  $g$  and  $f_j$  so that  $\partial\phi/\partial x$  exists for every nonnegative  $\mathbf{x}$ , then the conditions for the existence of a saddle point may be expressed in the form of differential inequalities.

► **THEOREM 7.1.3.** Given that  $X$  is the positive orthant, a necessary condition in order that  $\{\mathbf{x}^0, \mathbf{u}^0\}$  satisfy (7.1.2) for any continuously differentiable function  $\phi$  is that

$$\phi_x^0 = \left[ \frac{\partial \phi}{\partial x_i} \right]_{\{\mathbf{x}^0, \mathbf{u}^0\}} \leq 0, \quad (\phi_x^0, \mathbf{x}^0) = 0 \quad (\mathbf{x}^0 \geq 0) \quad (7.1.13)$$

and

$$\phi_u^0 = \left[ \frac{\partial \phi}{\partial u_j} \right]_{\{\mathbf{x}^0, \mathbf{u}^0\}} \geq 0, \quad (\phi_u^0, \mathbf{u}^0) = 0 \quad (\mathbf{u}^0 \geq 0). \quad (7.1.14)$$

If  $\phi(\mathbf{x}, \mathbf{u})$  is concave in  $\mathbf{x}$  and convex in  $\mathbf{u}$ , then (7.1.13) and (7.1.14) are both necessary and sufficient to establish (7.1.2).

*Proof. Necessity.* The relations (7.1.13) and (7.1.14) assert that  $\phi(\mathbf{x}, \mathbf{u}^0)$  has a local maximum at  $\mathbf{x} = \mathbf{x}^0$  and that  $\phi(\mathbf{x}^0, \mathbf{u})$  has a local minimum at  $\mathbf{u} = \mathbf{u}^0$ , respectively. This indeed must be the case if (7.1.2) holds.

*Sufficiency.* Since  $\phi(\mathbf{x}, \mathbf{u})$  is concave in  $\mathbf{x}$ , we have

$$\phi(\mathbf{x}, \mathbf{u}^0) - \phi(\mathbf{x}^0, \mathbf{u}^0) \leq (\phi_x^0, (\mathbf{x} - \mathbf{x}^0))$$

(property vii of Section B.4).

$$(\phi_x^0, (\mathbf{x} - \mathbf{x}^0)) = (\phi_x^0, \mathbf{x}) - (\phi_x^0, \mathbf{x}^0) \leq 0.$$

Hence

$$\phi(\mathbf{x}, \mathbf{u}^0) \leq \phi(\mathbf{x}^0, \mathbf{u}^0) \quad (\mathbf{x} \geq 0).$$

On the other hand, by the convexity of  $\phi(\mathbf{x}, \mathbf{u})$  in  $\mathbf{u}$  and by (7.1.14), we have

$$\phi(\mathbf{x}^0, \mathbf{u}) - \phi(\mathbf{x}^0, \mathbf{u}^0) \geq (\phi_u^0, \mathbf{u} - \mathbf{u}^0) = (\phi_u^0, \mathbf{u}) \geq 0 \quad (\mathbf{u} \geq 0).$$

**7.2 Examples of concave programming.** In this section we describe two models (portfolio selection and interindustry activity analysis) that lead naturally to concave programming problems. More extensive forms of nonlinear programming problems arise in connection with production and inventory studies.

In general there is no standard method available for solving nonlinear programming problems—i.e., nothing corresponding to the routine simplex algorithm of linear programming. Nevertheless a modest classification of types of nonlinear programming problems susceptible to a common approach can be accomplished. For example, as we have seen in Section 7.1, concave programming problems can be reduced to an equivalent Lagrangian saddle value form. Moreover, it is known and evident that the solution of a convex programming problem (maximizing a convex function where the variables range over a convex set  $X$ ) occurs at an extreme point of  $X$ .

Further particular qualitative results for special classes of problems may be obtained (see Chapter 8). It should be noted that since there is no single procedure for solving nonlinear programming problems—and this lack seems clearly inherent in the subject itself—the analysis of special examples is doubly important.

A. *Portfolio selection.* The process of selecting a portfolio of securities may be divided into two stages. The first stage is an appraisal of the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with a choice of portfolio. We are concerned here with the second stage. Specifically, we seek to determine what rule the investor should follow in order to maximize discounted expected future returns.

Let  $R_i$  represent the discounted return of one unit of the  $i$ th security. The value  $R_i$  can be further decomposed as an infinite sum of the form

$$R_i = \sum_{t=1}^{\infty} d_{it} r_{it},$$

where  $r_{it}$  is the anticipated return at time  $t$  per dollar invested in security  $i$ , and  $d_{it}$  is the rate of return on the  $i$ th security at time  $t$  discounted back to the present.

Suppose that only securities  $i = 1, 2, \dots, N$  are under consideration. Let  $x_i$  describe a given investment policy which denotes the proportions of security  $i$  bought; i.e.,

$$\sum_{i=1}^N x_i = 1.$$

We further assume  $x_i \geq 0$ , which implies that short sales are excluded as a possibility. The relative cumulative return for such a policy is

$$\sum_{i=1}^N R_i x_i = P(\mathbf{x}). \quad (7.2.1)$$

If we suppose further that  $R_i$  consists of normal random variables with known mean values  $\mu_i$  and covariance matrix  $\|\sigma_{ij}\|$ , then the mean value of  $P(\mathbf{x})$  is

$$\sum_{i=1}^N \mu_i x_i$$

and its variance is  $\sum_{i,j} x_i x_j \sigma_{ij}$ . Consider the objective function

$$g(\mathbf{x}) = a \sum_{i=1}^N \mu_i x_i - b \sum_{i,j} x_i x_j \sigma_{ij}, \quad (7.2.2)$$

where  $a$  and  $b$  are positive parameters whose relative magnitudes describe the importance of each of the terms of  $g(x)$ . The first is the expected relative yield according to the policy, which obviously is to be made large. The second is the measure of variability of the return, which can be regarded as the risk attending a given portfolio selection.

The problem, then, is to maximize  $g(x)$  subject to the constraints

$$x_i \geq 0, \quad \sum_{i=1}^N x_i = 1.$$

This is clearly a concave programming problem to which the results of Section 7.1 are applicable. The quadratic programming algorithm is also relevant (see Problem 13 and the notes to this chapter).

B. *An interindustry nonlinear programming model.* Suppose that for each of  $n$  desired commodities there is a production activity, an import activity, and an export activity. Let the variables of production, export, and import be

$X_j$  = amount of commodity  $j$  produced,

$M_j$  = amount of  $j$  imported,

$E_j$  = amount of  $j$  exported.

The parameters of the model are as follows:

$a_{ij}$  = input coefficient for commodity  $i$  used in production of  $j$   
(i.e., the amount of  $i$  used in producing one unit of  $j$ ),

$w_j$  = input coefficient for labor used in production of  $j$ ,

$c_j$  = input coefficient for capital used in production of  $j$ ,

$g_j$  = import unit price of  $j$ ,

$h_j$  = export unit price of  $j$ ,

$Y_j$  = final demand for  $j$  (which we shall assume to be strictly positive),

$L$  = available supply of labor,

$D$  = maximum allowable trade deficit.

(All the parameters of the model are naturally assumed to be nonnegative.)

The programming problem may be stated as follows: Minimize the total capital required,

$$c = \sum_j c_j X_j, \quad (7.2.3)$$

subject to the following restrictions:

Nonnegativity:

$$X_j \geq 0, \quad M_j \geq 0, \quad E_j \geq 0 \quad (j = 1, \dots, n). \quad (7.2.4)$$

Production:

$$X_j + M_j - E_j - \sum_{r=1}^n a_{jr} X_r \geq Y_j \quad (j = 1, \dots, n); \quad (7.2.5)$$

i.e., if  $E_j$  units are exported and

$$\sum_{r=1}^n a_{jr} X_r$$

units are used up by domestic production out of a total of  $X_j + M_j$  units produced and imported, enough should remain to satisfy the final demand  $Y_j$ .

Foreign exchange balance:

$$D + \sum_{j=1}^n h_j E_j - \sum_{j=1}^n g_j M_j \geq 0; \quad (7.2.6)$$

since the export revenue is

$$\sum_{j=1}^n h_j E_j$$

and the import cost is

$$\sum_{j=1}^n g_j M_j,$$

it follows that (7.2.6) expresses the allowable trade deficit.

Labor supply:

$$L - \sum_j w_j X_j \geq 0. \quad (7.2.7)$$

Two further natural assumptions are

$$a_{ij} \geq 0 \quad (i, j = 1, \dots, n), \quad \sum_i a_{ij} < 1 \quad (j = 1, \dots, n) \quad (7.2.8)$$

and

$$h_j = \gamma_j + \rho_j E_j, \quad \gamma_j > 0, \quad \rho_j < 0 \quad (j = 1, \dots, n). \quad (7.2.9)$$

The first assumption is simply that the output exceeds the input for every commodity, a necessary condition for efficient production plans. The second assumption is one commonly made in economics: to wit, the larger the amount of export goods made available, the smaller the return per unit. The linear functional form of (7.2.9) expresses a special case of this kind of dependence.

The reader will observe that in view of (7.2.9), the only nonlinearity in this model is contained in (7.2.6).

\**Formal statement of the method.* The function to be minimized, (7.2.3), is a convex function of  $X_j$  ( $j = 1, \dots, n$ ), and the set  $\{X, M, E\}$  satisfying restrictions (7.2.4-7) is a convex set. We shall now assume that there is a feasible  $\{X^0, M^0, E^0\}$  with strict inequality for (7.2.5), (7.2.6), and (7.2.7). Then, according to Theorem 7.1.1,

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{X} \\ \bar{M} \\ \bar{E} \end{pmatrix}$$

is optimal for the problem if and only if there exists a value

$$\bar{\mathbf{p}} = \begin{pmatrix} \bar{P} \\ \bar{P}_w \\ \bar{P}_f \end{pmatrix}$$

such that  $\{\bar{\mathbf{x}}, \bar{\mathbf{p}}\}$  is a saddle point of the Lagrangian form  $\phi(\mathbf{x}, \mathbf{p})$  defined by

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{p}) = & -\sum_j c_j X_j + \sum_j P_j (X_j - \sum_r a_{jr} X_r + M_j - E_j - Y_j) \\ & + P_w (L - \sum_j w_j X_j) + P_f (D - \sum_j g_j M_j + \sum_j h_j E_j); \end{aligned} \quad (7.2.10)$$

that is,

$$\phi(\mathbf{x}, \bar{\mathbf{p}}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{p}}) \leq \phi(\bar{\mathbf{x}}, \mathbf{p}) \quad (7.2.11)$$

$$\text{for all } \mathbf{x} = \begin{pmatrix} X \\ M \\ E \end{pmatrix} \geq \mathbf{0}, \quad \mathbf{p} = \begin{pmatrix} P \\ P_w \\ P_f \end{pmatrix} \geq \mathbf{0}.$$

It will be noted that by virtue of (7.2.5) and  $Y_j > 0$ , for any feasible

$$\begin{pmatrix} X \\ M \\ E \end{pmatrix}, \text{ we have}$$

$$X_j + M_j > 0 \quad (j = 1, \dots, n). \quad (7.2.12)$$

Substituting (7.2.9) into (7.2.10) and rearranging, we get

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{p}) = & P_w L + P_f D - \sum_j P_j Y_j + \sum_j (-c_j + P_j - \sum_i P_i a_{ij} - P_w w_j) X_j \\ & + \sum_j (P_j - P_f g_j) M_j + \sum_j [(-P_j + P_f \gamma_j) E_j + P_f \rho_j E_j^2]. \end{aligned} \quad (7.2.13)$$

---

\* The remainder of this section is of advanced nature and may be omitted at first reading without loss of continuity.

The saddle value problem will be solved in the following manner. We shall show that its  $\bar{\mathbf{p}}$  component is uniquely determined and may be explicitly written out. With  $\bar{\mathbf{p}}$  known, it is easy to secure the maximum  $\bar{\mathbf{x}}$  of  $\phi(\mathbf{x}, \bar{\mathbf{p}})$ , which completes the solution of the problem.

In order for

$$\mathbf{p} = \begin{pmatrix} P \\ P_w \\ P_f \end{pmatrix}$$

to be the  $\mathbf{p}$ -component of a saddle point, it is necessary (i) that

$$\max_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}, \mathbf{p})$$

be finite, and (ii) that there exist a value

$$\mathbf{x} = \begin{pmatrix} X \\ M \\ E \end{pmatrix}$$

which maximizes  $\phi(\mathbf{x}, \mathbf{p})$  and for which (7.2.12) is satisfied.

For the finiteness of (7.2.13) the coefficients of  $X_j$  and  $M_j$  must be non-positive. Thus (i) and (ii) are satisfied if and only if

$$P_j - P_f g_j \leq 0 \quad (7.2.14)$$

and

$$P_j - \sum_i P_i a_{ij} - P_w w_j - c_j \leq 0, \quad (7.2.15)$$

and for each  $j = 1, \dots, n$  equality holds for either (7.2.14) or (7.2.15) or both (since  $X_j + M_j > 0$ ).

We can replace conditions (7.2.14) and (7.2.15) by

$$P_j = \min \left\{ P_f g_j, \sum_i P_i a_{ij} + P_w w_j + c_j \right\} \quad (j = 1, \dots, n). \quad (7.2.16)$$

The relations (7.2.16) possess at most one solution  $P_j$  for prescribed values of  $P_f$  and  $P_w$ . To demonstrate this, we suppose that  $P_j$  and  $\tilde{P}_j$  ( $j = 1, \dots, n$ ) both satisfy (7.2.16) for given  $P_f$  and  $P_w$ . We consider the case

$$P_j = P_f g_j \leq \sum_i P_i a_{ij} + P_w w_j + c_j,$$

$$\tilde{P}_j = \sum_i \tilde{P}_i a_{ij} + P_w w_j + c_j \leq P_f g_j.$$

Then

$$|P_j - \tilde{P}_j| \leq \left| \sum_i (P_i - \tilde{P}_i) a_{ij} \right| \leq \left( \max_i |P_i - \tilde{P}_i| \right) c \quad (j = 1, \dots, n),$$

where  $c = \max_j \sum_i a_{ij} < 1$  by assumption (7.2.8). The same inequality ensues in all other circumstances, and we deduce that  $P_j = \bar{P}_j$  ( $j = 1, \dots, n$ ).

We now describe an iterative procedure for securing a solution of (7.2.16) with  $P_f$  and  $P_w$  specified.

The components  $(P_1, \dots, P_n)$  satisfying (7.2.16) will be determined in the order  $\dots, P_n^{(\nu)}, \dots, P_1^{(\nu)}, P_n^{(\nu+1)}, \dots, P_1^{(\nu+1)}, \dots$  from the recursive formulas

$$P_j^{(\nu+1)} = \min \left\{ P_f g_j, \sum_{i>j} P_i^{(\nu+1)} a_{ij} + \sum_{i \leq j} P_i^{(\nu)} a_{ij} + P_w w_j + c_j \right\} \quad (j = 1, \dots, n; \nu = 0, 1, 2, \dots), \quad (7.2.17)$$

with any nonnegative initial values  $P_1^{(0)}, \dots, P_n^{(0)}$ .

► **LEMMA 7.2.1.** For any nonnegative  $(P_w, P_f)$  and any nonnegative initial values  $(P_1^{(0)}, \dots, P_n^{(0)})$ , the values  $(P_1^{(\nu)}, \dots, P_n^{(\nu)})$  determined by (7.2.17) converge to a set of values  $(P_1, \dots, P_n)$  satisfying (7.2.16).

*Proof.* It will first be noted that

$$|P_j^{(\nu+1)} - P_j^{(\nu)}| \leq K c^\nu, \quad (7.2.18)$$

where  $K$  is a positive constant and  $c = \max_j \sum_i a_{ij}$ . Relation (7.2.18) is established by induction. Let us assume that

$$|P_j^{(k+1)} - P_j^{(k)}| \leq K c^k \quad (k = 0, 1, \dots, \nu - 1; j = 1, \dots, n; \\ k = \nu; j > i).$$

Then by (7.2.17) we have

$$\begin{aligned} |P_j^{(\nu+1)} - P_j^{(\nu)}| &\leq \left| \left( \sum_{i>j} P_i^{(\nu+1)} a_{ij} + \sum_{i \leq j} P_i^{(\nu)} a_{ij} + P_w w_j + c_j \right) \right. \\ &\quad \left. - \left( \sum_{i>j} P_i^{(\nu)} a_{ij} + \sum_{i \leq j} P_i^{(\nu-1)} a_{ij} + P_w w_j + c_j \right) \right| \\ &= \left| \sum_{i>j} (P_i^{(\nu+1)} - P_i^{(\nu)}) a_{ij} + \sum_{i \leq j} (P_i^{(\nu)} - P_i^{(\nu-1)}) a_{ij} \right| \\ &\leq \sum_{i>j} |P_i^{(\nu+1)} - P_i^{(\nu)}| a_{ij} + \sum_{i \leq j} |P_i^{(\nu)} - P_i^{(\nu-1)}| a_{ij} \\ &\leq K c^\nu \sum_{i>j} a_{ij} + K c^{\nu-1} \sum_{i \leq j} a_{ij} \\ &\leq K c^{\nu-1} \sum_i a_{ij} \leq K c^\nu, \end{aligned}$$

since  $c < 1$ .

The convergence of  $P_j^{(p)}$  to a  $P_j$  which obviously must satisfy (7.2.16) is now clear, and the proof of the lemma is complete.

For the triangular case, where

$$a_{ij} = 0, \quad i \leq j,$$

the recursive formulas (7.2.17) are replaced by the exact formula

$$P_j = \min \left\{ P_f g_j, \sum_{i>j} P_i a_{ij} + P_w w_j + c_j \right\}, \quad (7.2.19)$$

determined in the order  $j = n, \dots, 1$ .

A value

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{X} \\ \hat{M} \\ \hat{E} \end{pmatrix}$$

that maximizes  $\phi(\mathbf{x}, \mathbf{p})$  subject to  $\mathbf{x} \geq \mathbf{0}$  with  $\mathbf{p}$  fixed will be now determined. Let

$$S = \left\{ j; P_j < \sum_{i>j} P_i a_{ij} + P_w w_j + c_j \right\},$$

$$T = \{j; P_j < P_f g_j\}.$$

It follows from (7.2.19) that  $S$  and  $T$  are disjoint sets. Then

$$\hat{E}_j = \max \left\{ 0, \frac{P_f \gamma_j - P_j}{-2P_f \rho_j} \right\}, \quad (j = 1, \dots, n), \quad (7.2.20)$$

since the term involving  $E_j$  in (7.2.13) is a quadratic minimized by (7.2.20). Also, since the terms of  $X_j$  and  $M_j$  in (7.2.13) are linear it follows that  $\hat{X}_j = 0$  for  $j \in S$ , and  $\hat{M}_j = 0$  for  $j \in T$ . We may determine  $\hat{X}_j$  and  $\hat{M}_j$  as solutions to the following relations which maximize  $\phi(\mathbf{x}, \mathbf{p})$ . Explicitly,

$$\hat{X}_j - \sum_{k \notin S} a_{jk} \hat{X}_k = Y_j + \hat{E}_j \quad \text{for } j \notin S, \quad (7.2.21)$$

and

$$\hat{M}_j = Y_j + \hat{E}_j + \sum_{k \notin S} a_{jk} \hat{X}_k \quad \text{for } j \notin T. \quad (7.2.22)$$

With the aid of condition (7.2.8) it is easily proved that equations (7.2.21) possess a unique strictly positive solution. The values of  $\hat{X}_j$  and  $\hat{M}_j$ , both function of  $P_f$  and  $P_w$ , calculated by means of (7.2.21) and (7.2.22) satisfy the constraints (7.2.5).

Define

$$L(P_w, P_f) = \sum_j w_j \hat{X}_j$$

and

$$D(P_w, P_f) = \sum_j g_j \hat{M}_j - \sum_j \hat{h}_j \hat{E}_j,$$

where

$$\hat{h}_j = \gamma_j + \rho_j \hat{E}_j \quad (j = 1, \dots, n).$$

An application of Theorem 7.1.1 shows that

► **LEMMA 7.2.2.**  $(P_w, P_f)$  determines optimal  $(P_1, \dots, P_n, P_w, P_f)$  if and only if

$$L(P_w, P_f) \leq L \quad \text{and} \quad D(P_w, P_f) \leq D, \quad (7.2.23)$$

with equality when  $P_w > 0$  and  $P_f > 0$ , respectively.

Since  $L(P_w, P_f)$  and  $D(P_w, P_f)$  are approximately linear functions of  $P_w$  and  $P_f$ , optimal  $(P_w, P_f)$  can be approximated by linear interpolations.

If no choices of  $P_w$  and  $P_f$  suffice to satisfy (7.2.23), then (7.2.21) and (7.2.22) must be altered as follows:

$$\hat{X}_j + \hat{M}_j - \sum_R a_{jk} \hat{X}_k = Y_j + \hat{E}_j \quad \text{for } j \in R = \langle 1, 2, \dots, n \rangle - S - T,$$

$$\hat{X}_j - \sum_{k \notin S} a_{jk} \hat{X}_k = Y_j + \hat{E}_j \quad \text{for } j \in T,$$

and

$$\hat{M}_j = Y_j + \hat{E}_j + \sum_{k \notin S} a_{jk} \hat{X}_k \quad \text{for } j \in S - T.$$

The problem is then reduced to a standard problem of linear programming. In most cases, however, the solutions of (7.2.21) and (7.2.22) are valid.

We are interested in this model for two reasons. One is its usefulness in determining optimal plans for the production, export, and import of commodities. The second and more important reason is its usefulness as a tool for the investigation of similar nonlinear programming problems.

**\*7.3 The Arrow-Hurwicz gradient method.** We have seen by Theorems 7.1.1 and 7.1.2 that concave programming problems can be solved by finding a saddle point of the Lagrangian form  $\phi(\mathbf{x}, \mathbf{u})$ , where  $\phi(\mathbf{x}, \mathbf{u})$  is concave in  $\mathbf{x}$  and linear in  $\mathbf{u}$ . In this section we shall describe a method of approaching a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$  by solving a system of classical gradient equa-

\* Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

tions modified on the boundary to keep the variables in the positive orthant.

Let  $\phi(\mathbf{x}, \mathbf{u})$  be strictly concave in  $\mathbf{x} \geq \mathbf{0}$  and linear in  $\mathbf{u} \geq \mathbf{0}$ . For the first part of the theory it is enough to assume that  $\phi(\mathbf{x}, \mathbf{u})$  is convex in  $\mathbf{u}$ .

Consider the following system of differential equations:

$$\begin{aligned} \dot{x}_i &= \begin{cases} 0 & \text{if } x_i = 0 \text{ and } \phi_{x_i} < 0, \\ \phi_{x_i}[\mathbf{x}(t), \mathbf{u}(t)] & \text{otherwise,} \end{cases} \quad (i = 1, \dots, n), \\ \dot{u}_k &= \begin{cases} 0 & \text{if } u_k = 0 \text{ and } \phi_{u_k} > 0, \\ -\phi_{u_k}[\mathbf{x}(t), \mathbf{u}(t)] & \text{otherwise,} \end{cases} \quad (k = 1, \dots, m), \end{aligned} \quad (7.3.1)$$

where  $\phi_{x_i}$  represents partial differentiation with respect to the  $x_i$  variable and  $\phi_{u_k}$  has an analogous meaning. (Whenever we write  $\phi_{x_i}$  or  $\phi_{u_k}$  without indicating the points at which these derivatives are evaluated, the reader should interpret these derivatives in the obvious way.) In what follows we shall assume that the system (7.3.1) has a solution  $\{\mathbf{x}(t), \mathbf{u}(t)\}$  which is uniquely determined by any initial position  $\{\mathbf{x}^0, \mathbf{u}^0\} \geq \mathbf{0}$  and continuous with respect to  $\{\mathbf{x}^0, \mathbf{u}^0\}$ .

►THEOREM 7.3.1. Let  $\phi(\mathbf{x}, \mathbf{u})$  be strictly concave in  $\mathbf{x}$  and convex in  $\mathbf{u}$  and possess a saddle point. Then the solution  $\{\mathbf{x}(t), \mathbf{u}(t)\}$  of (7.3.1) with an arbitrary nonnegative initial position  $\{\mathbf{x}^0, \mathbf{u}^0\}$  converges to a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ .

*Proof.* We observe first that the  $\mathbf{x}$  component  $\bar{\mathbf{x}}$  of a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$  is uniquely determined. Assume to the contrary that  $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$  and  $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$  are two saddle points of  $\phi(\mathbf{x}, \mathbf{u})$  such that  $\bar{\mathbf{x}} \neq \bar{\mathbf{x}}$ ; then by the strict concavity of  $\phi$  in  $\mathbf{x}$ , and by the definition of a saddle point, we have

$$\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}) < \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}})$$

and

$$\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}) < \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}).$$

Hence

$$\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}) < \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}),$$

which is a contradiction.

Now consider the solution  $\{\mathbf{x}(t), \mathbf{u}(t)\}$  with an initial position  $\{\mathbf{x}^0, \mathbf{u}^0\}$ , and put

$$V(t) = \frac{1}{2} \{ |\mathbf{x}(t) - \bar{\mathbf{x}}|^2 + |\mathbf{u}(t) - \bar{\mathbf{u}}|^2 \},$$

where  $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ . (The term  $|\mathbf{x}(t) - \bar{\mathbf{x}}|$  represents

the Euclidean length of the vector  $\bar{\mathbf{x}}(t) - \bar{\mathbf{x}}$ , and  $|\mathbf{u}(t) - \bar{\mathbf{u}}|$  is similarly defined.)

We shall show that

$$\dot{V}(t) \leq 0 \quad \text{for any } t,$$

and

$$\dot{V}(t) < 0 \quad \text{for } t \text{ such that } \mathbf{x}(t) \neq \bar{\mathbf{x}}.$$

We observe that

$$\begin{aligned} \dot{V}(t) &= (\dot{\mathbf{x}}(t), \mathbf{x}(t) - \bar{\mathbf{x}}) + (\dot{\mathbf{u}}(t), \mathbf{u}(t) - \bar{\mathbf{u}}) \\ &= (\phi_{x(t)}, \mathbf{x}_{I(t)} - \bar{\mathbf{x}}_{I(t)}) - (\phi_{u(t)}, \mathbf{u}_{I(t)} - \bar{\mathbf{u}}_{I(t)}), \end{aligned} \quad (7.3.2)$$

where

$$\mathbf{x}_{I(t)} = \delta_{x(t)} \cdot \mathbf{x}(t), \quad \mathbf{u}_{I(t)} = \delta_{u(t)} \cdot \mathbf{u}(t),$$

$$\bar{\mathbf{x}}_{I(t)} = \delta_{x(t)} \cdot \bar{\mathbf{x}}, \quad \bar{\mathbf{u}}_{I(t)} = \delta_{u(t)} \cdot \bar{\mathbf{u}},$$

$$\delta_{x(t)} = \begin{vmatrix} \delta_{x_1(t)} & & 0 \\ & \ddots & \\ 0 & & \delta_{x_n(t)} \end{vmatrix}, \quad \delta_{u(t)} = \begin{vmatrix} \delta_{u_1(t)} & & 0 \\ & \ddots & \\ 0 & & \delta_{u_m(t)} \end{vmatrix},$$

$$\delta_{x_i(t)} = \begin{cases} 0 & \text{if } x_i(t) = 0 \text{ and } \phi_{x_i(t)} < 0, \\ 1 & \text{otherwise,} \end{cases}$$

$$\delta_{u_k(t)} = \begin{cases} 0 & \text{if } u_k(t) = 0 \text{ and } \phi_{u_k(t)} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Since  $\phi(\mathbf{x}, \mathbf{u})$  is strictly concave in  $\mathbf{x}$  and convex in  $\mathbf{u}$ , we have

$$\phi[\bar{\mathbf{x}}, \mathbf{u}(t)] - \phi[\mathbf{x}(t), \mathbf{u}(t)] < (\phi_{x(t)}, \bar{\mathbf{x}} - \mathbf{x}(t)) \quad \text{for } \bar{\mathbf{x}} \neq \mathbf{x}(t), \quad (7.3.3)$$

and

$$\phi[\mathbf{x}(t), \bar{\mathbf{u}}] - \phi[\mathbf{x}(t), \mathbf{u}(t)] \geq (\phi_{u(t)}, \bar{\mathbf{u}} - \mathbf{u}(t)) \quad (7.3.4)$$

(see Section B.4). Moreover, since  $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{u})$ , we have

$$\phi[\mathbf{x}(t), \bar{\mathbf{u}}] \leq \phi[\bar{\mathbf{x}}, \bar{\mathbf{u}}] \leq \phi[\bar{\mathbf{x}}, \mathbf{u}(t)].$$

Hence

$$0 \leq \phi[\bar{\mathbf{x}}, \mathbf{u}(t)] - \phi[\mathbf{x}(t), \bar{\mathbf{u}}] < (\phi_{x(t)}, \bar{\mathbf{x}} - \mathbf{x}(t)) - (\phi_{u(t)}, \bar{\mathbf{u}} - \mathbf{u}(t)). \quad (7.3.5)$$

Set

$$\mathbf{x}(t) = \mathbf{x}_{I(t)} + \mathbf{x}_{II(t)}, \quad \mathbf{u}(t) = \mathbf{u}_{I(t)} + \mathbf{u}_{II(t)}.$$

From (7.3.2) and (7.3.5), after noticing that

$$(\phi_{x(t)}, \mathbf{x}_{II(t)}(t)) = 0, \quad (\phi_{u(t)}, \mathbf{u}_{II(t)}(t)) = 0,$$

we obtain

$$\begin{aligned} \dot{V}(t) &= (\phi_{x(t)}, \mathbf{x}(t) - \bar{\mathbf{x}}) - (\phi_{x(t)}, \mathbf{x}_{II(t)} - \bar{\mathbf{x}}_{II(t)}) \\ &\quad - (\phi_{u(t)}, \mathbf{u}(t) - \bar{\mathbf{u}}) + (\phi_{u(t)}, \mathbf{u}_{II(t)} - \bar{\mathbf{u}}_{II(t)}) \\ &< (\phi_{x(t)}, \bar{\mathbf{x}}_{II(t)}) - (\phi_{u(t)}, \bar{\mathbf{u}}_{II(t)}) \leq 0, \quad \text{if } \mathbf{x}(t) \neq \bar{\mathbf{x}}. \end{aligned} \quad (7.3.6)$$

The last inequality is valid, since by definition  $\bar{\mathbf{x}}_{II(t)}$  has its only nonzero components where  $\phi_{x(t)} < 0$ , and  $\bar{\mathbf{u}}_{II(t)}$  has its only nonzero components where  $\phi_{u(t)} > 0$ . Similarly, we have  $\dot{V}(t) \leq 0$  for any  $t$ .

Now we shall prove that the  $\mathbf{x}$  component  $\mathbf{x}(t)$  converges to the uniquely determined  $\mathbf{x}$  component  $\bar{\mathbf{x}}$  of saddle points.

Since  $\dot{V}(t) \leq 0$  for any  $t$  and  $V(t) \geq 0$ , there exists a value

$$V^* = \lim_{t \rightarrow \infty} V(t).$$

Hence the set  $\{\mathbf{x}(t) | t \geq 0\}$  is bounded. Let  $\mathbf{x}^*$  be any accumulation point of  $\{\mathbf{x}(t) | t \geq 0\}$ . Then there is a sequence  $\{t_\nu\}$  such that

$$\lim_{\nu \rightarrow \infty} \mathbf{x}(t_\nu) = \mathbf{x}^*$$

and

$$\lim_{\nu \rightarrow \infty} \mathbf{u}(t_\nu) = \mathbf{u}^*,$$

for a suitable  $\mathbf{u}^*$ . By the assumption of uniqueness the solution with initial position  $\{\mathbf{x}(t_\nu), \mathbf{u}(t_\nu)\}$  is  $\{\mathbf{x}(t + t_\nu), \mathbf{u}(t + t_\nu)\}$ . Since the solution depends continuously on its initial position, we obtain that

$$\{\mathbf{x}^*(t), \mathbf{u}^*(t)\} = \lim_{\nu \rightarrow \infty} \{\mathbf{x}(t + t_\nu), \mathbf{u}(t + t_\nu)\},$$

where  $\{\mathbf{x}^*(t), \mathbf{u}^*(t)\}$  represents the solution with starting position  $\{\mathbf{x}^*, \mathbf{u}^*\}$ . But

$$\begin{aligned} V^* &= \lim_{\nu \rightarrow \infty} \frac{1}{2} \{|\mathbf{x}(t + t_\nu) - \bar{\mathbf{x}}|^2 + |\mathbf{u}(t + t_\nu) - \bar{\mathbf{u}}|^2\} \\ &= \frac{1}{2} \{|\mathbf{x}^*(t) - \bar{\mathbf{x}}|^2 + |\mathbf{u}^*(t) - \bar{\mathbf{u}}|^2\}. \end{aligned} \quad (7.3.7)$$

This is incompatible with (7.3.6) unless  $\mathbf{x}^* = \bar{\mathbf{x}}$ . Consequently, every accumulation point of  $\{\mathbf{x}(t)\}$  is  $\bar{\mathbf{x}}$  and we conclude that

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \bar{\mathbf{x}}.$$

Lastly, we shall prove that the  $\mathbf{u}$  component  $\mathbf{u}(t)$  of the solution converges to the  $\mathbf{u}$  component  $\tilde{\mathbf{u}}$  of a saddle point  $\{\bar{\mathbf{x}}, \tilde{\mathbf{u}}\}$ , although  $\tilde{\mathbf{u}}$  need not coincide with  $\bar{\mathbf{u}}$ . Since  $V(t) \rightarrow V^*$  and  $\mathbf{x}(t) \rightarrow \bar{\mathbf{x}}$ , we find that  $\mathbf{u}(t)$  converges and denote its limit by  $\tilde{\mathbf{u}}$ . Since  $\phi_u(\bar{\mathbf{x}}, \mathbf{u})$  is independent of  $\mathbf{u}$  ( $\phi$  is linear in  $\mathbf{u}$ ), we infer that

$$\phi_u(\bar{\mathbf{x}}, \tilde{\mathbf{u}}) = \phi_u(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \geq 0, \quad (7.3.8)$$

the second relation being valid because  $\{\bar{\mathbf{x}}, \bar{\mathbf{u}}\}$  is a saddle point (Theorem 7.1.3). But  $\dot{V}(t)$  tends to zero, and hence by (7.3.2)

$$(\phi_{x(t)}, \mathbf{x}_{I(t)} - \bar{\mathbf{x}}_{I(t)}) = (\phi_u(t), \mathbf{u}_{I(t)} - \bar{\mathbf{u}}_{I(t)})$$

converges to zero. However, since  $\mathbf{x}_{I(t)} \rightarrow \bar{\mathbf{x}}_{I(t)}$ , it follows that

$$(\phi_u(\bar{\mathbf{x}}, \tilde{\mathbf{u}}), \tilde{\mathbf{u}}_I - \bar{\mathbf{u}}_I) = 0$$

or

$$(\phi_u(\bar{\mathbf{x}}, \tilde{\mathbf{u}}), \tilde{\mathbf{u}}_I) = (\phi_u(\bar{\mathbf{x}}, \bar{\mathbf{u}}), \bar{\mathbf{u}}_I),$$

again exploiting the fact that  $\phi_u(\bar{\mathbf{x}}, \mathbf{u})$  is independent of  $\mathbf{u}$ . The indices of  $\tilde{\mathbf{u}}_I$  include the terms in which  $\tilde{u}_j > 0$  by the definition of  $I$ . However,  $\phi_u(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \geq 0$ ,  $\bar{\mathbf{u}} \geq \mathbf{0}$ , and  $(\phi_u(\bar{\mathbf{x}}, \bar{\mathbf{u}}), \bar{\mathbf{u}}) = 0$  by Theorem 7.1.3. Therefore, by (7.3.8) and the definition of  $\tilde{\mathbf{u}}_I$ , we have

$$0 \leq (\phi_u(\bar{\mathbf{x}}, \tilde{\mathbf{u}}), \tilde{\mathbf{u}}) \leq (\phi_u(\bar{\mathbf{x}}, \bar{\mathbf{u}}), \bar{\mathbf{u}}) = 0,$$

which gives us

$$(\phi_u(\bar{\mathbf{x}}, \tilde{\mathbf{u}}), \tilde{\mathbf{u}}) = 0. \quad (7.3.9)$$

Comparing (7.3.8) and (7.3.9) with the formulas of Theorem 7.1.3, we conclude that  $\{\bar{\mathbf{x}}, \tilde{\mathbf{u}}\}$  is a saddle point as asserted.

**7.4 The vector maximum problem.** The nonlinear programming problem of Section 7.1 may be extended to deal with the problem of maximizing a vector function  $\mathbf{G}(\mathbf{x})$  in  $E^r$  subject to the constraints  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ ,  $\mathbf{x} \in X$ , where  $X$  is closed and convex. The components of  $\mathbf{G}(\mathbf{x})$  and  $\mathbf{F}(\mathbf{x})$  are assumed to be concave functions of  $\mathbf{x}$ . An  $\mathbf{x}^0$  satisfying the constraints (i.e., a feasible vector) is said to be an “efficient point” if there exists no other vector  $\mathbf{x}$  satisfying the constraints such that  $\mathbf{G}(\mathbf{x}) > \mathbf{G}(\mathbf{x}^0)$ . The vector maximum problem is solved when all efficient points have been determined.

The constraints  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$  and  $\mathbf{x} \in X$  as before are assumed to satisfy the conditions of Theorem 7.1.1. An efficient point  $\mathbf{x}^0$  satisfying the constraints is characterized by constructing an ordinary maximization problem with the same constraints such that  $\mathbf{x}^0$  is a maximum point. For this purpose, the following lemma is fundamental.

► **LEMMA 7.4.1.** If  $\mathbf{x}^0$  is an efficient point, then there exists a vector  $\mathbf{v} = \langle v_1, \dots, v_r \rangle$  with  $v_i \geq 0$  and

$$\sum_{i=1}^r v_i = 1$$

such that the maximum of the function  $g(\mathbf{x}) = (\mathbf{v}, \mathbf{G}(\mathbf{x}))$  in the set of all  $\mathbf{x}$  satisfying  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ ,  $\mathbf{x} \in X$  is achieved for  $\mathbf{x} = \mathbf{x}^0$ .

*Proof.* The constraint set, which is clearly convex, will be denoted by  $S$ . Let

$$H = \{\mathbf{y} = \mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{x}^0) \in E^r \mid \mathbf{x} \in S\},$$

and let  $K$  be the convex set spanned by  $H$ . Since  $\mathbf{x}^0$  is efficient, the set  $H$  cannot overlap with the interior of the positive orthant in  $E^r$ . It follows immediately from the concavity of  $G$  and the fact that  $\mathbf{x}^0$  is efficient that the set  $K$  does not contain an interior point of the positive orthant. Let

$$\xi = \langle \xi_1^0, \xi_2^0, \dots, \xi_r^0, c \rangle \quad (\xi \neq 0)$$

represent a hyperplane which separates the positive orthant from  $K$ . The arguments of Lemma B.1.3 apply here, and we see that  $\xi_i^0 \geq 0$  and  $c = 0$ . Set

$$v_i^0 = \frac{\xi_i^0}{\sum_j \xi_j^0}.$$

Then it is readily verified that the vector  $\mathbf{v}^0$  satisfies the assertion of the lemma.

*Remark 7.4.1.* Let us assume momentarily that  $G$  has continuous partial derivatives. If the set  $H$  is not tangent to any coordinate direction at zero, then the separating hyperplane  $\xi$  may be chosen so that it touches the first orthant only at zero. In this case, we see that  $v_i^0 > 0$  for every  $i$ . Some conditions are now cited which guarantee the above property for  $H$ .

► **COROLLARY 7.4.1.** If  $\mathbf{x}^0$  is an efficient interior point of the constraint set  $S$  with the property that  $\mathbf{G}_{x^0} \cdot d\mathbf{x} \geq \mathbf{0}$  for no vector differential  $d\mathbf{x}$ , then the corresponding vector  $\mathbf{v}$  of Lemma 7.4.1 may be chosen possessing all positive components.

The same conclusion holds if  $\mathbf{x}^0$  is efficient and belongs to the boundary of the constraint set, provided  $\mathbf{G}_{x^0} \cdot d\mathbf{x} \geq \mathbf{0}$  for no vector differential satisfying  $\mathbf{F}_{x^0} d\mathbf{x} \geq \mathbf{0}$ ,  $\mathbf{I}_1 d\mathbf{x} \geq \mathbf{0}$ , where the inequalities  $\mathbf{F}(\mathbf{x}^0) \geq \mathbf{0}$  and  $\mathbf{I}\mathbf{x}^0 \geq \mathbf{0}$  ( $I$  being the identity matrix) have been separated into the two sets of conditions  $\mathbf{F}_1(\mathbf{x}^0) = \mathbf{0}$ ,  $\mathbf{I}_1\mathbf{x}^0 = \mathbf{0}$  and  $\mathbf{F}_2(\mathbf{x}^0) > \mathbf{0}$ ,  $\mathbf{I}_2\mathbf{x}^0 > \mathbf{0}$ .

The saddle point problem that is equivalent to the efficient point problem can now be given.

► **THEOREM 7.4.1.** Let  $\mathbf{G}$  be a concave vector function defined for  $\mathbf{x} \in X$ , and let  $\mathbf{F}$  be the same as in Theorem 7.1.1. In order for  $\mathbf{x}^0$  to be an efficient point, it is necessary that there exist two nonnegative vectors  $\mathbf{v}^0$  ( $\mathbf{v}^0 \neq \mathbf{0}$ ) and  $\mathbf{u}^0$  with the same number of components as  $\mathbf{G}$  and  $\mathbf{F}$ , respectively, such that  $\{\mathbf{x}^0, \mathbf{u}^0\}$  is a saddle point of

$$\chi(\mathbf{x}, \mathbf{u}, \mathbf{v}^0) = (\mathbf{v}^0, \mathbf{G}(\mathbf{x})) + (\mathbf{u}, \mathbf{F}(\mathbf{x}));$$

that is,

$$\chi(\mathbf{x}, \mathbf{u}^0, \mathbf{v}^0) \leq \chi(\mathbf{x}^0, \mathbf{u}^0, \mathbf{v}^0) \leq \chi(\mathbf{x}^0, \mathbf{u}, \mathbf{v}^0) \quad (7.4.1)$$

for all nonnegative  $\mathbf{u}$  and  $\mathbf{x} \in X$ . If  $\mathbf{v}^0$  has all components strictly positive, then (7.4.1) is also sufficient.

*Proof. Necessity.* By Lemma 7.4.1 there exists a nonnegative vector  $\mathbf{v}^0$  such that the maximum of  $g(\mathbf{x}) = (\mathbf{v}^0, \mathbf{G}(\mathbf{x}))$  in the set  $S$  is achieved for  $\mathbf{x}^0$ . The conclusion of (7.4.1) is obtained by applying Theorem 7.1.1.

*Sufficiency.* Reversing the reasoning, it follows again by Theorem 7.1.1 that the maximum of  $(\mathbf{v}^0, \mathbf{G}(\mathbf{x}))$  in  $S$  is achieved for  $\mathbf{x}^0$ . Since  $\mathbf{v}^0$  is a strictly positive vector, it also follows that  $\mathbf{x}^0$  is efficient.

The economic interpretation of the saddle-point problem (7.4.1) is analogous to that of (7.1.2). Since the desired output is now represented not by a real number but by a vector, the concept of maximum is replaced by that of efficient point.

**\*7.5 Conjugate functions.** A further extension of the results of linear programming to nonlinear maximization and minimization problems involving concave and convex functions can be effected by means of the theory of conjugate functions as set forth by Fenchel. Aside from its value in nonlinear programming, the theory is elegant and interesting in itself and has additional applications to the theory of inequalities and to the theory of the structure of convex sets. Both these topics embrace many of the fundamental working tools of finite game theory and programming problems. In the following sections we develop the basic properties of conjugate functions. Their relevance to nonlinear programming and convex set theory is described in Sections 7.8 and 7.9, respectively.

Let  $f$  be a convex function defined on a convex set  $C$  in  $E^n$ . Recall that  $f$  is continuous in the relative interior of  $C$ , and that if  $\mathbf{x}^*$  is on the boundary of  $C$  and  $f(\mathbf{x}^*)$  is defined, then

$$\liminf_{x \rightarrow x^*} f(\mathbf{x}) \leq f(\mathbf{x}^*)$$

(see Section B.4).

For the discussion that follows, whenever

$$\lim_{x \rightarrow x^*} f(x)$$

is finite, the set  $C$  is redefined to include all points  $x^*$ . Likewise, the definition of  $f$  is altered at such points so that

$$f(x^*) = \lim_{x \rightarrow x^*} f(x).$$

This modified function and the enlarged set will be denoted by  $f^0$  and  $C^0$ , respectively.

► **DEFINITION 7.5.1.** A convex function defined on a convex set  $S$  will be called closed if for any  $x^0$  where

$$\lim_{x \rightarrow x^0} f(x)$$

is finite,  $f(x^0)$  is defined and is equal to

$$\lim_{x \rightarrow x^0} f(x).$$

The process of obtaining  $f^0$  and  $C^0$  from  $f$  and  $C$  will be referred to as “closing the function,” and  $[f^0, C^0]$  will be called the “closure” of  $[f, C]$ . The closure of  $[f^0, C^0]$  is again  $[f^0, C^0]$ .

► **LEMMA 7.5.1.** If  $f$  is a convex function defined on a convex set  $C$  in  $E^n$ , then its closure  $f^0$  is convex with domain of definition  $C^0$ .

*Proof.* Let  $x^* = t\mathbf{x}_1 + (1 - t)\mathbf{x}_2$ , with  $0 \leq t \leq 1$  and  $\mathbf{x}_1, \mathbf{x}_2 \in C^0$ . We can find sequences of interior points  $\mathbf{x}_1^n \rightarrow \mathbf{x}_1$  and  $\mathbf{x}_2^n \rightarrow \mathbf{x}_2$  such that

$$\lim_{n \rightarrow \infty} f^0(\mathbf{x}_1^n) = \lim_{n \rightarrow \infty} f(\mathbf{x}_1^n) = f^0(\mathbf{x}_1)$$

and

$$\lim_{n \rightarrow \infty} f(\mathbf{x}_2^n) = f^0(\mathbf{x}_2).$$

Then

$$\begin{aligned} f^0(\mathbf{x}^*) &= \lim_{x \rightarrow x^*} f(x) \leq \lim_{n \rightarrow \infty} f[t\mathbf{x}_1^n + (1 - t)\mathbf{x}_2^n] \leq t f^0(\mathbf{x}_1) \\ &\quad + (1 - t) f^0(\mathbf{x}_2). \end{aligned}$$

Henceforth we shall assume that the original convex function  $f$  defined on  $C$  is already closed.

We now construct the set

$$[f, C] = \{ \{z, \mathbf{x}\} \in E^{n+1} \mid \mathbf{x} \in C \text{ and } z \geq f(\mathbf{x}) \}. \quad (7.5.1)$$

The lower boundary of this set is the graph of the function  $f$ . Since  $f$  is assumed to be closed and convex, it follows readily that  $[f, C]$  is a closed convex set in  $E^{n+1}$ .

A convex set  $S$  of points in  $E^n$  can also be characterized by specifying the set of all its supporting hyperplanes (see Section B.1). It is therefore natural in the study of  $S$  to consider the set of all its supporting planes (determined by a normal unit vector and a real number) as a dual set  $S^*$  in  $E^n$  space. Starting with  $S^*$  and constructing its dual set leads essentially back to  $S$ . We now explore the consequences of this duality principle as applied to  $[f, C]$ . The *conjugate set*  $[f, C]^*$  is defined as the set of all nonvertical hyperplanes in  $E^{n+1}$  with the property that the set  $[f, C]$  lies above the plane. More precisely, a plane in  $E^{n+1}$  is determined by a vector  $\xi$  in  $E^n$  and a real number  $c$  such that  $-c$  is the intercept of the plane and the vertical axis of  $E^{n+1}$ . The point  $\{c, \xi\}$  in  $E^{n+1}$  belongs to  $[f, C]^*$  if and only if

$$z \geq (\mathbf{x}, \xi) - c \quad \text{for every } \{z, \mathbf{x}\} \text{ in } [f, C]. \quad (7.5.2)$$

In order to verify that a point  $\{c, \xi\}$  is a member of  $[f, C]^*$ , it is enough to satisfy the inequalities

$$c \geq (\mathbf{x}, \xi) - f(\mathbf{x}) \quad \text{for every } \mathbf{x} \text{ in } C. \quad (7.5.3)$$

Clearly, if  $\{c_1, \xi\}$  is in  $[f, C]^*$  and  $c_2 > c_1$ , then  $\{c_2, \xi\}$  is in  $[f, C]^*$ . Consequently, for a given vector  $\xi$ , the smallest admissible  $c$  is

$$\sup_{\mathbf{x} \in C} [(\mathbf{x}, \xi) - f(\mathbf{x})]. \quad (7.5.4)$$

Let

$$\Gamma = \{\xi \text{ in } E^n \mid \sup_{\mathbf{x} \in C} [(\mathbf{x}, \xi) - f(\mathbf{x})] < \infty\}$$

and set

$$\phi(\xi) = \sup_{\mathbf{x} \in C} [(\mathbf{x}, \xi) - f(\mathbf{x})]$$

for  $\xi$  in  $\Gamma$ . The function  $\phi$  is referred to as the conjugate function of  $f$ . It is the largest vertical intercept of planes with normal  $\{-1, \xi\}$  lying below the graph of  $f$ .

► **LEMMA 7.5.2.** The set  $\Gamma$  is convex and nonvoid, and  $\phi$  is a convex closed function with domain  $\Gamma$ .

*Proof.* If  $C$  has an interior in  $E^n$ , we construct a supporting hyperplane to  $[f, C]$  at  $\mathbf{x}^0$ , an interior point of  $C$ . Such a plane is clearly nonvertical. If  $C$  has no interior in  $E^n$ , let  $\mathbf{x}^0$  denote a relative interior point of  $C$  and  $\bar{C}$  the linear extension of  $C$  in  $E^n$ . We may consider  $f$  defined in the affine†

---

† An affine space is a translation of a linear space.

space  $\widetilde{C} \supset C$ . A supporting plane to  $[f, C]$  at  $\mathbf{x}^0$ , defined only in the smaller space of the values of  $f$  and the set  $\widetilde{C}$ , is a nonvertical hyperplane which may be extended in infinitely many ways to  $E^n$  while remaining a nonvertical hyperplane. Thus in any case  $[\phi, \Gamma]$  is nonvoid. The proof that  $\phi$  is convex is straightforward.

We indicate the proof that  $\phi$  is also closed. Let  $\xi^n \rightarrow \xi^*$ ; then  $\phi(\xi^n) \geq (\mathbf{x}, \xi^n) - f(\mathbf{x})$ , and hence

$$\varliminf_n \phi(\xi^n) \geq \lim_n (\mathbf{x}, \xi^n) - f(\mathbf{x}) = (\mathbf{x}, \xi^*) - f(\mathbf{x})$$

for all  $\mathbf{x} \in C$ . Therefore

$$\varliminf_n \phi(\xi^n) \geq \sup_{\mathbf{x} \in C} [(\mathbf{x}, \xi^*) - f(\mathbf{x})] = \phi(\xi^*).$$

Since the opposite inequality always holds for convex functions, we must have equality.

*Remark 7.5.1.* The discussion preceding this last lemma implies that

$$[f, C]^* = [\phi, \Gamma]. \quad (7.5.5)$$

The striking conclusion of Lemma 7.5.2 and (7.5.5) is that the dual set to  $[f, C]$  is again of the same type. Boundary points of  $[\phi, \Gamma]$  correspond to points on the graph of  $\phi(\xi)$  and may be viewed as planes of support to the graph of  $f(\mathbf{x})$ . If  $\{f(\mathbf{x}), \mathbf{x}\}$  is a point of contact between the set  $[f, C]$  and the plane  $\{\phi(\xi), \xi\}$  with normal vector of components  $\{-1, \xi\}$  and intercept  $\phi(\xi)$  in the vertical direction, then  $f(\mathbf{x}) + \phi(\xi) = (\mathbf{x}, \xi)$ . The next lemma expresses the duality principle common to the study of convex sets.

►LEMMA 7.5.3.  $[[f, C]^*]^* = [f, C]$ .

*Proof.* We first show that  $[f, C] \subset [\phi, \Gamma]^* = [[f, C]^*]^*$ . Let  $\mathbf{x}$  be in  $C$ ; then  $\phi(\xi) \geq (\mathbf{x}, \xi) - f(\mathbf{x})$  for any  $\xi$  in  $\Gamma$ . But  $z \geq (\mathbf{x}, \xi) - \phi(\xi)$  whenever  $z \geq f(\mathbf{x})$  for all  $\xi$  in  $\Gamma$ . Therefore  $\{z, \mathbf{x}\}$  is in  $[\phi, \Gamma]^*$ .

In order to establish the converse, let  $\{z^0, \mathbf{x}^0\}$  be a point in  $E^{n+1}$  but not in  $[f, C]$ . Since  $[f, C]$  is closed and convex, there exists a nonvertical hyperplane that separates the point  $\{z^0, \mathbf{x}^0\}$  from the set  $[f, C]$ . Let the plane be described by  $\{c, \xi\}$ . Then  $f(\mathbf{x}) \geq (\mathbf{x}, \xi) - c$  for all  $\mathbf{x}$  in  $C$ , and  $z^0 < (\mathbf{x}^0, \xi) - c$ . The first inequality implies that  $\{c, \xi\}$  is in  $[\phi, \Gamma]$ , and the second that  $\{z^0, \mathbf{x}^0\}$  is not in  $[\phi, \Gamma]^*$ . Hence  $[f, C] \supset [\phi, \Gamma]^*$ , which together with the first result proves the lemma.

Since  $f(\mathbf{x})$  is conjugate to  $\phi(\xi)$ ,

$$f(\mathbf{x}) = \sup_{\xi \in \Gamma} [(\mathbf{x}, \xi) - \phi(\xi)]. \quad (7.5.6)$$

*Boundary points.* If  $\xi^0$  is in  $\Gamma$ , then  $\{\phi(\xi^0), \xi^0\}$  is a point on the boundary of  $[\phi, \Gamma]$  and represents a supporting plane to  $[f, C]$ . Let the contact set be  $C(\xi^0)$ . Any point  $\{f(\mathbf{x}^0), \mathbf{x}^0\}$  in the contact set is a supporting hyperplane to  $[\phi, \Gamma]$  at the point  $\{\phi(\xi^0), \xi^0\}$ . Thus if  $\{f(\mathbf{x}^0), \mathbf{x}^0\}$  belongs to  $C(\xi^0)$ , then  $\{\phi(\xi^0), \xi^0\}$  is in the intersection of  $[\phi, \Gamma]$  and the plane  $\{f(\mathbf{x}^0), \mathbf{x}^0\}$ . If the contact sets corresponding to  $\mathbf{x}^0$  and  $\xi^0$  consist of unique interior points, and if partial derivatives exist for both  $f(\mathbf{x})$  and  $\phi(\xi)$  at  $\mathbf{x}^0$  and  $\xi^0$ , respectively, then, since  $\phi(\xi^0) + f(\mathbf{x}^0) = (\mathbf{x}^0, \xi^0)$ , we have

$$\left( \frac{\partial f}{\partial x_i} \right)_{x^0} = \xi_i^0, \quad \left( \frac{\partial \phi}{\partial \xi_i} \right)_{\xi^0} = x_i^0. \quad (7.5.7)$$

**\*7.6 Composition of conjugate functions.** Three operations are commonly applied to convex sets in building new convex sets. They are: (1) forming direct sums of convex sets, (2) constructing the convex hulls of sets, and (3) taking the intersections of convex sets (see Appendix B). In this section, we investigate the application of these operations to the conjugate sets introduced above.

Let  $[f_1, C_1]$  and  $[f_2, C_2]$  be two closed convex sets of the type described in (7.5.1). Let their conjugate sets be  $[\phi_1, \Gamma_1]$  and  $[\phi_2, \Gamma_2]$ , respectively. The vector sum  $S = [\phi_1, \Gamma_1] + [\phi_2, \Gamma_2]$  is the set of all points of the form  $\{a_1 + a_2, \xi^1 + \xi^2\}$ , where  $\{a_1, \xi^1\}$  traverses  $[\phi_1, \Gamma_1]$  and  $\{a_2, \xi^2\}$  traverses  $[\phi_2, \Gamma_2]$ . The closure of  $S$  will be denoted by  $\bar{S}$ .

► **THEOREM 7.6.1.** If  $C = C_1 \cap C_2$  is nonvoid and  $f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$  for  $\mathbf{x}$  in  $C$ , then  $[f, C]^* = \bar{S} = [\phi, \Gamma]$ , where  $\Gamma_1 + \Gamma_2 \subset \Gamma \subset \overline{\Gamma_1 + \Gamma_2}$  and

$$\phi(\xi) = \inf_{\substack{\xi^1 \in \Gamma_1, \xi^2 \in \Gamma_2 \\ \xi^1 + \xi^2 = \xi}} \{ \phi_1(\xi^1) + \phi_2(\xi^2) \} \quad (7.6.1)$$

for  $\xi$  in  $\Gamma_1 + \Gamma_2$ .

*Proof.* For clarity, the proof is divided into three separate parts.

$$(i) \quad [f, C]^* \supset \bar{S}.$$

The conjugate of  $[f, C]$  is the set of all nonvertical planes  $\{a, \xi\}$  for which  $a \geq (\mathbf{x}, \xi) - f(\mathbf{x})$  for all  $\mathbf{x}$  in  $C$ . If  $\{a_i, \xi^i\}$  belongs to  $[\phi_i, \Gamma_i]$ , then  $a_i \geq (\xi^i, \mathbf{x}) - f_i(\mathbf{x})$  for  $i = 1, 2$ . By adding, we obtain that  $a_1 + a_2 \geq (\xi^1 + \xi^2, \mathbf{x}) - f(\mathbf{x})$  for  $\mathbf{x}$  in  $C$ , so that  $\{a_1 + a_2, \xi^1 + \xi^2\}$  is in  $[f, C]^*$ . Hence  $[f, C]^* \supset \bar{S}$ , and since  $[f, C]^*$  is closed, it follows that  $[f, C]^* \supset \bar{S}$ .

$$(ii) \quad \phi(\xi) \quad \text{is convex and} \quad \bar{S} = [\phi, \Gamma].$$

Next we define  $\phi(\xi)$  for  $\xi$  in  $\Gamma_1 + \Gamma_2$  according to (7.6.1). In view of (i),  $\phi(\xi)$  is finite for each  $\xi$  in  $\Gamma_1 + \Gamma_2$ . From its definition, we may verify that

$\phi(\xi)$  is convex. Moreover, by examining the construction of  $\bar{S}$ , we see that  $\bar{S} = [\phi, \Gamma]$  as asserted.

$$(iii) \quad \bar{S} \supset [f, C]^*.$$

It suffices to verify that  $\bar{S}^* = [\phi, \Gamma]^* \subset [f, C]$ . By definition the set  $\bar{S}^*$  is composed of all nonvertical barriers to  $S$ ; hence its elements can be represented in the form  $\{z, \mathbf{x}\}$ . A point  $\{z, \mathbf{x}\}$  is in  $\bar{S}^*$  if and only if for every  $\{a_1, \xi^1\}$  and  $\{a_2, \xi^2\}$  in  $[\phi_1, \Gamma_1]$  and  $[\phi_2, \Gamma_2]$ , respectively,

$$z \geq (\mathbf{x}, \xi^1 + \xi^2) - a_1 - a_2.$$

Since  $a_1 \geq \phi_1(\xi^1)$  and  $a_2 \geq \phi_2(\xi^2)$ , this last condition is equivalent to  $z \geq (\mathbf{x}, \xi^1 + \xi^2) - \phi_1(\xi^1) - \phi_2(\xi^2)$ . Hence

$$z \geq \sup_{\xi^1 \in \Gamma^1} [(\mathbf{x}, \xi^1) - \phi_1(\xi^1)] + \sup_{\xi^2 \in \Gamma^2} [(\mathbf{x}, \xi^2) - \phi_2(\xi^2)] = f_1(\mathbf{x}) + f_2(\mathbf{x}).$$

Consequently,  $\{z, x\}$  is in  $[f, C]$ . This applies to every point in  $\bar{S}^*$ , and the proof of the last statement is therefore complete.

Let  $[f_\alpha, C_\alpha]$  be a family of closed convex sets in  $E^{n+1}$  of the form given in (7.5.1). Let  $C \subset \cap_\alpha C_\alpha$  be the set of all  $x$  for which  $\sup_\alpha f_\alpha(x)$  is finite. On  $C$ , define  $f(\mathbf{x}) = \sup_\alpha f_\alpha(\mathbf{x})$ ; then  $C$  is convex and  $f(\mathbf{x})$  is a convex closed function defined over  $C$ . In terms of the point set  $[f, C] = \cap_\alpha [f_\alpha, C_\alpha]$ , we wish to characterize the set  $[\phi, \Gamma] = [f, C]^*$ . Let  $[\phi_\alpha, \Gamma_\alpha] = [f_\alpha, C_\alpha]^*$  and let  $H$  be the convex hull of  $\{\Gamma_\alpha\}$ .

► **THEOREM 7.6.2.** If  $C$  is nonvoid and  $f(\mathbf{x}) = \sup_\alpha f_\alpha(\mathbf{x})$  for  $\mathbf{x}$  in  $C$ , then  $[f, C]^*$  is the closure of the convex hull  $G$  of  $[\phi_\alpha, \Gamma_\alpha]$ . Furthermore,  $\bar{G} = [\phi, \Gamma]$ ,  $H \subset \Gamma \subset \bar{H}$ ; and

$$\phi(\xi) = \inf_{\sum \lambda_i \xi^{\alpha_i} = \xi} \sum_{i=1}^{n+1} \lambda_i \phi_{\alpha_i}(\xi^{\alpha_i}), \quad (7.6.2)$$

the infimum operation being extended over all representations of  $\xi$  as a centroid of  $n+1$  points taken from any  $n+1$  of the sets  $\Gamma_\alpha$ .

*Proof.* Since  $[f, C] \subset [f_\alpha, C_\alpha]$ , it follows that  $[f, C]^* \supset [f_\alpha, C_\alpha]^* = [\phi_\alpha, \Gamma_\alpha]$ . The fact that  $[f, C]^*$  is closed and convex implies that it must contain all finite convex combinations of  $[\phi_\alpha, \Gamma_\alpha]$  and their limits. Hence  $[f, C]^* \supset \bar{G}$ .

Since  $G$  in  $E^{n+1}$  is the convex hull of  $[\phi_\alpha, \Gamma_\alpha]$ , we may define  $\phi(\xi)$  as the smallest ordinate (i.e., first component) with  $\{\phi(\xi), \xi\}$  in  $\bar{G}$ . Since  $[\phi_\alpha, \Gamma_\alpha]$  spans  $G$  convexly in  $E^{n+1}$ , every point of  $\bar{G}$  can be approached by points of the form

$$\xi = \sum_{i=1}^{n+2} \lambda_i \xi^{\alpha_i} \quad \text{and} \quad c = \sum_{i=1}^{n+2} \lambda_i c_{\alpha_i}$$

with  $\{c_{\alpha_i}, \xi^{\alpha_i}\}$  in  $[\phi_{\alpha_i}, \Gamma_{\alpha_i}]$ . (See Lemma B.2.1.) But  $\phi(\xi)$  is the smallest admissible  $c$  for the vector  $\xi$ . Therefore

$$\phi(\xi) = \inf_{\sum \lambda_i \xi^{\alpha_i} = \xi} \sum_{i=1}^{n+2} \lambda_i \phi_{\alpha_i}(\xi^{\alpha_i}).$$

The point  $\{\phi(\xi), \xi\}$  is on the boundary of  $\overline{G}$ ; it is therefore located in a simplex of dimension at most  $n$ . Since extreme points of  $\overline{G}$  must arise from extreme points of  $[\phi_{\alpha_i}, \Gamma_{\alpha_i}]$ , the corners of this simplex can be made to correspond to points  $\{\phi_{\alpha_i}(\xi^{\alpha_i}), \xi^{\alpha_i}\}$ . Hence, in evaluating  $\phi(\xi)$  it suffices to consider convex combinations of at most  $n+1$  points. With the definition of  $\phi(\xi)$  as proposed, we may easily verify that  $\phi(\xi)$  is a closed function defined on  $\Gamma$  as indicated and that  $\overline{G} = [\phi, \Gamma]$ .

It remains to prove that  $[f, C]^* = \overline{G}$ . In view of the preceding, we need only establish that  $[f, C]^* \subset \overline{G} = [\phi, \Gamma]$ , or equivalently  $[f, C] \supset [\phi, \Gamma]^*$ . Let  $\{z^0, \mathbf{x}^0\}$  be a point of  $[\phi, \Gamma]^*$ . Then for every  $\alpha$  and  $\xi^\alpha$  in  $\Gamma_\alpha$ , we have  $z^0 \geq (\xi^\alpha, \mathbf{x}^0) - \phi_\alpha(\xi^\alpha)$ . Hence  $\{z^0, \mathbf{x}^0\}$  is in  $[\phi_\alpha, \Gamma_\alpha]^*$  for every  $\alpha$ , and therefore  $\{z^0, \mathbf{x}^0\} \in \cap_\alpha [f_\alpha, C_\alpha] = [f, C]$ .

Another useful consequence of the definition of conjugate functions is the following theorem.

► **THEOREM 7.6.3.** Let  $f(\mathbf{x})$  in  $C$  be a closed convex function and  $\phi(\xi)$  its conjugate function. Then the conjugate function of  $f(\mathbf{x}) + k$  in  $C$ , where  $k$  is a constant, is  $\phi(\xi) - k$  in  $\Gamma$ . The conjugate of  $f(\mathbf{x} - \mathbf{v})$  in  $C + \mathbf{v}$ , where  $\mathbf{v}$  is a constant vector, is  $\phi(\xi) + (\mathbf{v}, \xi)$  in  $\Gamma$ .

Our next theorem will be of value in the theory of convex games as well as in the study of convex sets (see Volume II, Section 4.3).

► **THEOREM 7.6.4.** Let  $f_\alpha(\mathbf{x})$  be a family of convex functions defined over a bounded convex set  $C$  in  $E^n$ , such that  $\sup_\alpha f_\alpha(\mathbf{x})$  is finite for each  $\mathbf{x}$ . Then for any  $a \leq \inf_x \sup_\alpha f_\alpha(\mathbf{x})$  and  $\epsilon > 0$ , there exist functions  $f_{\alpha_i}$  and constants  $\lambda_i \geq 0$  ( $i = 1, \dots, n+1$ ) such that

$$\sum_{i=1}^{n+1} \lambda_i f_{\alpha_i}(\mathbf{x}) \geq a - \epsilon \quad \text{for all } \mathbf{x} \in C$$

and  $\sum \lambda_i = 1$ .

*Proof.* Let  $f(\mathbf{x}) = \sup_\alpha f_\alpha(\mathbf{x})$ , and  $z_0 = \inf_x f(\mathbf{x})$ . The horizontal plane of height  $z_0$  is a supporting plane to the convex set  $[f, C]$ , and it is also a

boundary point of  $[f, C]^*$ . For this plane,  $\xi = \mathbf{0}$  (the zero vector) and  $\phi(\mathbf{0}) = -z_0$ . It now follows from Theorem 7.6.2 that

$$-z_0 = \phi(\mathbf{0}) = \inf_{\sum \lambda_i \xi^{\alpha_i} = \mathbf{0}} \sum_{i=1}^{n+1} \lambda_i \phi_{\alpha_i}(\xi^{\alpha_i}),$$

where  $\lambda_i \geq 0$  and  $\sum \lambda_i = 1$ . Therefore, for the given  $\epsilon > 0$ , we can find  $\lambda_i$  and  $\xi^{\alpha_i}$  such that

$$-z_0 \geq \sum_{i=1}^{n+1} \lambda_i \phi_{\alpha_i}(\xi^{\alpha_i}) - \epsilon \quad \text{and} \quad \sum_{i=1}^{n+1} \lambda_i \xi^{\alpha_i} = \mathbf{0}.$$

But for any  $\mathbf{x} \in C$ , we have  $\phi_{\alpha_i}(\xi^{\alpha_i}) \geq (\mathbf{x}, \xi^{\alpha_i}) - f_{\alpha_i}(\mathbf{x})$ . Then

$$z_0 \leq - \sum_{i=1}^{n+1} \lambda_i (\mathbf{x}, \xi^{\alpha_i}) + \sum_{i=1}^{n+1} \lambda_i f_{\alpha_i}(\mathbf{x}) + \epsilon = \sum_{i=1}^{n+1} \lambda_i f_{\alpha_i}(\mathbf{x}) + \epsilon.$$

Since  $a - \epsilon \leq z_0 - \epsilon$ , the theorem is valid.

*Support functions.* Given any convex set  $C \subset E^n$ , we may define  $f(\mathbf{x}) \equiv 0$  on  $C$ . The dual set to  $[f, C]$  is of the form  $[\phi, \Gamma]$ , where

$$\phi(\xi) = \sup_{\mathbf{x} \in C} (\mathbf{x}, \xi)$$

and the domain  $\Gamma$  is the set of vectors for which

$$\sup_{\mathbf{x} \in C} (\mathbf{x}, \xi)$$

is finite. We call  $\phi(\xi)$  the support function of the set  $C$ .

►LEMMA 7.6.1. The support function  $\phi(\xi)$  of a convex set  $C$  is a positively homogeneous function, and  $[\phi, \Gamma]$  is a cone with vertex at the origin. Conversely, if a function  $\phi$  defined over a closed convex cone of  $E^n$  with vertex at the origin is positively homogeneous, then  $\phi(\xi)$  is the support function of some convex set  $C$  in  $E^n$ .

*Proof.* If  $k > 0$ , then

$$\phi(k\xi) = \sup_{\mathbf{x} \in C} (\mathbf{x}, k\xi) = k \sup_{\mathbf{x} \in C} (\mathbf{x}, \xi) = k\phi(\xi).$$

Hence  $\Gamma$  and  $[\phi, \Gamma]$  are cones with vertex at the origin.

Conversely, let  $\phi(\xi)$  be a positively homogeneous function defined over a closed convex cone  $\Gamma$  with vertex at the origin. Then  $[\phi, \Gamma]$  is a cone of the same type, and the extreme points of  $[\phi, \Gamma]^*$  must be planes through

the origin. Hence  $[\phi, \Gamma]^* = [f, C]$  where  $f \equiv 0$  on  $C$ , and  $C$  is convex. Finally,

$$[\phi, \Gamma] = [\phi, \Gamma]^{**} = [f, C]^*,$$

so that  $\phi$  is the support function of the set  $C$ .

The representation of the support function of an intersection of convex sets in terms of the support functions of the individual sets can be deduced from (7.6.2). Let  $C$  be a family of convex sets in  $E^n$ , and let  $\phi$  be their support functions. If  $C = \bigcap_{\alpha} C_{\alpha}$  and  $\phi$  is the support function of  $C$ , then by Theorem 7.6.2

$$\begin{aligned}\phi(\xi) &= \inf_{\sum \lambda_i \xi^{\alpha_i} = \xi} \sum_{i=1}^{n+1} \lambda_i \phi_{\alpha_i}(\xi^{\alpha_i}) = \inf_{\sum \lambda_i \xi^{\alpha_i} = \xi} \sum_{i=1}^{n+1} \phi_{\alpha_i}(\lambda_i \xi^{\alpha_i}) \\ &= \inf_{\sum \eta^{\alpha_i} = \xi} \sum_{i=1}^{n+1} \phi_{\alpha_i}(\eta^{\alpha_i}),\end{aligned}$$

by the positive homogeneity of  $\phi_{\alpha_i}(\xi)$ .

**\*7.7 Conjugate concave functions.** A conjugate function theory for concave functions  $g(\mathbf{x})$  defined over a convex set  $D$  can be developed analogously to the theory for conjugate convex functions. We define the convex set

$$[g, D] = (\{z, \mathbf{x}\} \mid \mathbf{x} \in D \quad \text{and} \quad z \leq g(\mathbf{x})).$$

By an appropriate redefinition of  $g(\mathbf{x})$  and  $D$ , we can make the set  $[g, D]$  closed (see Section 7.5). We may then define the conjugate set of  $[g, D]$  in terms of nonvertical planes which keep  $[g, D]$  on one side. A nonvertical plane is determined by a vector  $\xi$  and a number  $b$ , where  $-b$  is the  $z$  intercept of the plane. The barrier plane is in  $[g, D]^*$  if and only if  $(\xi, \mathbf{x}) - b \geq g(\mathbf{x})$  or

$$b \leq (\xi, \mathbf{x}) - g(\mathbf{x}). \quad (7.7.1)$$

If  $b_1$  is admissible for a vector  $\xi$  and if  $b_2 < b_1$ , then  $b_2$  is also admissible by (7.7.1). If we denote by  $\psi(\xi)$  the largest admissible  $b$ , then

$$\psi(\xi) = \inf_{\mathbf{x} \in C} [(\xi, \mathbf{x}) - g(\mathbf{x})]. \quad (7.7.2)$$

The domain  $\Delta$  of  $\psi$  is the set of vectors for which (7.7.2) is finite. It is easy to verify that  $\Delta$  is convex,  $\psi$  is concave, and  $[\psi, \Delta]$  is the conjugate of  $[g, D]$ . Furthermore,  $[\psi, \Delta]^* = [g, D]$  and

$$g(\mathbf{x}) = \inf_{\xi \in \Delta} [(\mathbf{x}, \xi) - \psi(\xi)]. \quad (7.7.3)$$

All of the algebraic and geometric characteristics of the composition of conjugate concave functions carry over. In particular, the analog of Theorem 7.6.4 becomes:

► **THEOREM 7.7.1.** Let  $g_\alpha(\mathbf{x})$  be a family of concave functions defined over a bounded convex set  $C$  in  $E^n$ , such that  $\inf_\alpha g_\alpha(\mathbf{x})$  is finite for each  $\mathbf{x}$ . Then for any  $a \geq \sup_x \inf_\alpha g_\alpha(\mathbf{x})$  and  $\epsilon > 0$ , there exist functions  $g_{\alpha_i}$  and constants  $\lambda_i \geq 0$  ( $i = 1, \dots, n+1$ ) such that

$$\sum_{i=1}^{n+1} \lambda_i g_{\alpha_i}(\mathbf{x}) \leq a + \epsilon \quad \text{for all } \mathbf{x} \in C$$

and  $\sum \lambda_i = 1$ .

\***7.8 A duality theorem of nonlinear programming.** Let  $f(\mathbf{x})$  be convex over a convex set  $C$  and  $g(\mathbf{x})$  be concave over a convex set  $D$ . Let  $[f, C]^* = [\phi, \Gamma]$  and  $[g, D]^* = [\psi, \Delta]$ .

*Hypothesis.* We assume that  $C \cap D$  and  $\Gamma \cap \Delta$  have points which are relatively interior to  $C, D$  and  $\Gamma, \Delta$ , respectively.

Consider the following extremum problems.

*Problem I.* Find a point  $\mathbf{x}^0$  in  $C \cap D$  that maximizes  $g(\mathbf{x}) - f(\mathbf{x})$  in  $C \cap D$ .

If  $g(\mathbf{x}) - f(\mathbf{x}) \geq 0$  in  $C \cap D$ , this problem, stated geometrically, is to find the maximal vertical chord of the convex set  $[f, C] \cap [g, D]$  in  $E^{n+1}$ . If  $f(\mathbf{x}) \equiv 0$  in  $C$ , this problem reduces to the programming problem, i.e., the problem of maximizing  $g(\mathbf{x})$  for  $\mathbf{x}$  in  $C$ .

*Problem II.* Find a point  $\xi^0$  in  $\Gamma \cap \Delta$  that minimizes  $\phi(\xi) - \psi(\xi)$  in  $\Gamma \cap \Delta$ .

If  $\phi(\xi) - \psi(\xi) \geq 0$  in  $\Gamma \cap \Delta$ , this problem, stated geometrically, is to find the minimum vertical segment joining the sets  $[\phi, \Gamma]$  and  $[\psi, \Delta]$  in  $E^{n+1}$ .

These two problems may be considered as extensions of Problems I and II of linear programming to the case of nonlinear programming. However, the problems here are not equivalent to the saddle-value problem discussed in Section 7.1. In some situations the dual problem can be given an economic or industrial interpretation.

► **THEOREM 7.8.1.**  $\sup_{x \in C \cap D} [g(\mathbf{x}) - f(\mathbf{x})] = \inf_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)]$ .

*Proof.* Let  $\xi \in \Gamma \cap \Delta$ . Then

$$\begin{aligned} \phi(\xi) &\geq (\mathbf{x}, \xi) - f(\mathbf{x}) & (\mathbf{x} \in C), \\ \psi(\xi) &\leq (\mathbf{x}, \xi) - g(\mathbf{x}) & (\mathbf{x} \in D), \\ \phi(\xi) - \psi(\xi) &\geq g(\mathbf{x}) - f(\mathbf{x}) & (\mathbf{x} \in C \cap D). \end{aligned}$$

Hence

$$\inf_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)] \geq \sup_{x \in C \cap D} [g(x) - f(x)]. \quad (7.8.1)$$

Relation (7.8.1) shows that both quantities are finite.

In order to prove the opposite inequality, we set

$$\mu = \sup_{x \in C \cap D} [g(x) - f(x)]$$

and  $f_1(x) = f(x) + \mu$ . The conjugate of  $f_1(x)$  will be denoted by  $\phi_1(\xi)$ . The choice of  $\mu$  allows us to deduce that the sets  $[f_1, C]$  and  $[g, D]$  come arbitrarily close to each other but do not overlap at any interior points. Therefore there exists a separating plane  $h$  which is a supporting plane to both sets. We distinguish two cases.

*Case 1.* If  $h$  is not a vertical plane, then  $h$  is determined by a vector  $\xi^0$  and its  $z$  intercept  $-\eta$ . Consequently  $\xi^0$  belongs to  $\Gamma \cap \Delta$ ,  $\eta = \psi(\xi^0)$ , and

$$\eta = \phi_1(\xi^0) = \sup_{x \in C} [(x, \xi^0) - f(x) - \mu] = \phi(\xi^0) - \mu.$$

Hence

$$\sup_{x \in C \cap D} [g(x) - f(x)] = \mu = \phi(\xi^0) - \psi(\xi^0) \geq \inf_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)],$$

and equality holds in (7.8.1).

*Case 2.* If  $h$  is a vertical plane, we project  $[f, C]$ ,  $[g, D]$ , and  $h$  vertically onto  $E^n$ . The two convex sets project onto  $C$  and  $D$ , and the projection of  $h$  is a hyperplane  $h_1$  of  $E^n$  which must separate  $C$  and  $D$ . The last assertion contradicts the hypothesis that  $C$  and  $D$  have relative interior points in common. Thus Case 2 is impossible.

► **COROLLARY 7.8.1.** The proof of the last theorem actually shows that

$$\inf_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)] = \phi(\xi^0) - \psi(\xi^0),$$

with  $\xi^0$  in  $\Gamma \cap \Delta$ . Therefore we may replace “infimum” by “minimum.” Since the hypothesis of the theorem is symmetric in  $f, g$  and  $\phi, \psi$ , we can also replace “supremum” by “maximum.” Thus

$$\min_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)] = \max_{x \in C \cap D} [g(x) - f(x)]. \quad (7.8.2)$$

► **COROLLARY 7.8.2.** If we replace  $g(x)$  and  $\phi(\xi)$  by their equivalents in terms of  $\psi$  and  $f$ , respectively,

$$g(x) = \inf_{\xi \in \Delta} [(x, \xi) - \psi(\xi)], \quad \phi(\xi) = \sup_{x \in C} [(x, \xi) - f(x)], \quad (7.8.3)$$

the theorem asserts that

$$\max_{x \in C \cap D} \inf_{\xi \in \Delta} [(\mathbf{x}, \xi) - f(\mathbf{x}) - \psi(\xi)] = \min_{\xi \in \Gamma \cap \Delta} \sup_{x \in C} [(\mathbf{x}, \xi) - f(\mathbf{x}) - \psi(\xi)].$$

If the given sets and functions have the additional properties that  $\Gamma = \Delta$ ,  $C = D$ , and  $\inf, \sup$  may be replaced by  $\min, \max$  in (7.8.3), then

$$\max_{x \in C} \min_{\xi \in \Delta} [(\mathbf{x}, \xi) - f(\mathbf{x}) - \psi(\xi)] = \min_{\xi \in \Delta} \max_{x \in C} [(\mathbf{x}, \xi) - f(\mathbf{x}) - \psi(\xi)]. \quad (7.8.4)$$

This equation amounts to demonstrating the existence of a saddle point for the function  $F(\mathbf{x}, \xi) = (\mathbf{x}, \xi) - f(\mathbf{x}) - \psi(\xi)$ . Therefore, there exist  $\xi^0$  and  $\mathbf{x}^0$  such that

$$F(\mathbf{x}^0, \xi) \geq F(\mathbf{x}^0, \xi^0) \geq F(\mathbf{x}, \xi^0), \quad \text{for all } \mathbf{x} \in C \text{ and } \xi \in \Delta. \quad (7.8.5)$$

Conversely, whenever (7.8.5) is satisfied, Theorem 7.8.1 is valid.

If the expressions for  $\psi$  and  $g$  are substituted into (7.8.2), then, analogously to (7.8.4), we obtain

$$\max_{x \in C} \min_{\xi \in \Delta} [\phi(\xi) + g(\mathbf{x}) - (\mathbf{x}, \xi)] = \min_{\xi \in \Delta} \sup_{x \in C} [\phi(\xi) + g(\mathbf{x}) - (\mathbf{x}, \xi)]. \quad (7.8.6)$$

This amounts to demonstrating the existence of a saddle point for the function  $\phi(\xi, \mathbf{x}) = \phi(\xi) + g(\mathbf{x}) - (\mathbf{x}, \xi)$ .

A stronger form of Theorem 7.8.1 can be obtained which involves no special assumptions about the existence of relative interior points to  $C \cap D$  and  $\Gamma \cap \Delta$ . All that is required is that both these last sets be nonvoid. The proof is based on the nature of the composition of conjugate functions. Let  $\psi(\xi)$  be the conjugate of  $f(\mathbf{x}) + [-g(\mathbf{x})]$  in  $C \cap D$ . It follows from Theorem 7.6.1 that  $\psi(\xi)$  is defined on a set containing  $\Gamma + (-\Delta)$ . Since  $\Gamma$  and  $\Delta$  have common points,  $\Gamma - \Delta$  contains the origin. Hence  $\psi(\mathbf{0})$  is defined and by Theorem 7.6.1

$$\begin{aligned} \psi(\mathbf{0}) &= \inf_{\substack{\xi^1 \in \Gamma, \xi^2 \in -\Delta \\ \xi^1 + \xi^2 = 0}} [\phi(\xi^1) - \psi(-\xi^2)] \\ &= \inf_{\xi \in \Gamma \cap \Delta} [\phi(\xi) - \psi(\xi)]. \end{aligned}$$

On the other hand, the very definition of the conjugate of  $f(\mathbf{x}) - g(\mathbf{x})$ , when taken for  $\xi = \mathbf{0}$ , yields

$$\psi(\mathbf{0}) = \sup_{x \in C \cap D} [g(\mathbf{x}) - f(\mathbf{x})].$$

This last argument suffers from the defect that it does not indicate when  $\inf, \sup$  can be replaced by  $\min, \max$  (see Corollary 7.8.1).

**\*7.9 Applications of the theory of conjugate functions to convex sets.** Theorem 7.6.4 points the way to a series of results about convex sets, approximation to function, fitting polynomials to prescribed loci, etc. These structural properties of convex sets are of constant use in the study of inequalities occurring in various aspects of programming problems. The following classical theorem, which expresses a strengthened form of the finite intersection property characterizing compactness, is fundamental.

► **THEOREM 7.9.1. (Helly's Theorem).** If  $C_\alpha$  is a collection of bounded, closed, convex sets in  $E^n$  such that every  $n + 1$  sets intersect, then  $\cap_\alpha C_\alpha$  is nonvoid.

*Proof.* Let  $\phi_\alpha(\mathbf{y})$  be the function measuring the distance between  $\mathbf{y}$  and  $C_\alpha$  in  $E^n$ . It is trivial to verify that each  $\phi_\alpha$  is continuous and convex. Let  $S$  be a closed bounded sphere containing one of the  $C_\alpha$ , say  $C_{\alpha_0}$ . If  $\cap_\alpha C_\alpha$  is void, then

$$\inf_{\mathbf{y} \in S} \sup_\alpha \phi_\alpha(\mathbf{y}) > 0.$$

Since  $\phi_\alpha$  are all continuous functions, we can find a finite subfamily  $\{C_{\alpha_i}\}$  such that

$$\inf_{\mathbf{y} \in S} \sup_i \phi_{\alpha_i}(\mathbf{y}) > 0.$$

Since  $S$  is compact, we may write

$$\inf_{\mathbf{y} \in S} \sup_i \phi_{\alpha_i}(\mathbf{y}) \geq \delta > 0.$$

It follows from Theorem 7.6.4 that there exists a convex combination (which for convenience of notation we label as the first  $n + 1$ ) satisfying

$$\sum_{i=1}^{n+1} \lambda_i \phi_{\alpha_i}(\mathbf{y}) \geq \delta > 0 \quad (\mathbf{y} \in S). \quad (*)$$

Therefore the  $n + 1$  sets  $C_{\alpha_i}$  must intersect outside  $S$ .

Let  $S'$  be a closed bounded sphere containing  $S$ ,  $C_{\alpha_0}$ , and the  $C_{\alpha_i}$  ( $i = 1, \dots, n + 1$ ) of equation (\*). Then

$$\inf_{\mathbf{y} \in S'} \sup_i \phi_{\alpha_i}(\mathbf{y}) > 0.$$

Repeating the previous argument for the family consisting of the  $n + 2$  elements  $C_{\alpha_i}$  ( $i = 0, \dots, n + 1$ ), we conclude that some  $n + 1$  of them do not intersect over  $S'$ . Since they are all contained in  $S'$ , they do not intersect at all, contradicting the hypothesis.

►COROLLARY 7.9.1. If every  $n + 1$  sets of the family  $C_\alpha$  possesses a nonvoid common intersection, and if either

- (a)  $C_\alpha$  is a finite family of open convex sets (not necessarily bounded), or
- (b)  $C_\alpha$  is a finite family of closed convex sets (not necessarily bounded), or
- (c)  $C_\alpha$  is an infinite family of closed convex sets (not necessarily bounded, but possessing a finite subfamily with a bounded intersection),

then  $\cap_\alpha C_\alpha$  is nonvoid.

In each case modifying the sets appropriately yields an equivalent problem to which Theorem 7.9.1 is applicable.

The theorem is not true for an arbitrary family of closed convex sets. In  $E^1$ , let  $C_k = [k, \infty]$  ( $k = 1, 2, \dots$ ). Any finite number of sets intersect, but there is no common point for the whole family.

An important application of Helly's Theorem is to the Polynomial Approximation Problem. Given  $n + 1$  pairs of real numbers

$$\langle x_i, y_i \rangle \quad (i = 0, \dots, n; x_i < x_{i+1}),$$

we seek to find the polynomial  $P$  of degree at most  $n - 1$  that minimizes the quantity  $\max_i |y_i - P(x_i)|$ . Let

$$\rho = \min_P \max_i |y_i - P(x_i)|.$$

That  $\rho$  is actually achieved by a polynomial  $P^*$  is easily proved by a routine compactness argument. We leave the details for the reader. Later, an explicit expression for  $\rho$  and  $P^*$  is noted.

►LEMMA 7.9.1. If  $P^*$  is a solution, then  $|y_i - P^*(x_i)| = \rho$  for all  $i$ .

*Proof.* (By contradiction.) Assume that there exists some  $i$ , say  $i = 0$ , such that  $|y_0 - P^*(x_0)| < \rho$ . We shall show that this assumption would allow us to construct a better polynomial approximation, contradicting the choice of  $\rho$ .

For each  $k \neq 0$  we may construct a polynomial  $Q_k(x)$  of degree  $n - 1$  such that  $Q_k(x_i) = 0$  except for  $i = 0$  and  $i = k$ . Furthermore, we can always choose  $Q_k(x)$  such that  $Q_k(x_k) > 0$ .

Let  $\epsilon_k = \text{sign}[y_k - P^*(x_k)]$ . Select  $a_i > 0$  ( $i = 1, 2, \dots, n$ ) so small that

$$|y_0 - P^*(x_0) - a_i \epsilon_i Q_i(x_0)| < \rho.$$

This is possible in view of the supposition that  $|y_0 - P^*(x_0)| < \rho$ . But  $a_i > 0$  and  $|y_i - P^*(x_i)| \leq \rho$  imply that

$$|y_i - P^*(x_i) - a_i \epsilon_i Q_i(x_i)| < \rho, \quad (7.9.1)$$

while the values at the remaining  $x_j$  are not disturbed because  $Q_i$  vanishes there. Hence

$$\left| y_i - P^*(x_i) - \sum_{k=1}^n a_k \epsilon_k Q_k(x_i) \right| < \rho \quad \text{for all } i. \quad (7.9.2)$$

But (7.9.2) is in contradiction to the definition of  $\rho$ .

►LEMMA 7.9.2. If  $P^*(x)$  is a solution, then the  $y_i - P^*(x_i)$  alternate in sign.

*Proof.* If for some  $x_i$  and  $x_{i+1}$ , the values  $y_i - P^*(x_i)$  are of the same sign, say positive, we can construct a polynomial of degree  $n - 1$  that vanishes at the remaining  $x_i$  and is positive at  $x_i$  and at  $x_{i+1}$ . By subtracting an appropriate multiple of this polynomial we can reduce the error at  $x_i$  and  $x_{i+1}$  without disturbing the errors at the other points. This contradicts the assertion of Lemma 7.9.1.

It is clear from Lemmas 7.9.1 and 7.9.2 that if

$$P^*(x) = \sum_{i=0}^{n-1} a_i^* x^i$$

is a solution, it must satisfy the system of equations

$$(-1)^k \eta + \sum_{i=0}^{n-1} a_i^* x_k^i = y_k \quad (k = 0, \dots, n), \quad (7.9.3)$$

where  $\eta$  is either  $+\rho$  or  $-\rho$ .

The solution is actually unique. If two solutions exist with the same value of  $\eta$ , the difference  $P_1^* - P_2^*$  must be a polynomial of degree at most  $n - 1$  vanishing at  $n + 1$  points  $x_0, \dots, x_n$ . Hence  $P_1^* = P_2^*$ . On the other hand, if two solutions exist such that  $\eta_1 = +\rho$  and  $\eta_2 = -\rho$ , their difference must vanish at some point between  $x_i$  and  $x_{i+1}$  for each  $i$ . Hence it vanishes for at least  $n$  points, and this means it is zero.

We may regard (7.9.3) as a system of  $n + 1$  equations in the  $n + 1$  unknowns  $a_0^*, \dots, a_n^*$ . The value of  $\eta$  is determined uniquely by

$$\eta = \frac{\begin{vmatrix} y_0 & 1 & x_0 & \dots & x_0^{n-1} \\ y_1 & 1 & x_1 & \dots & x_1^{n-1} \\ \vdots & & & & \\ y_n & 1 & x_n & \dots & x_n^{n-1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & x_0 & \dots & x_0^{n-1} \\ -1 & 1 & x_1 & \dots & x_1^{n-1} \\ \vdots & & & & \\ (-1)^n & 1 & x_n & \dots & x_n^{n-1} \end{vmatrix}}.$$

The coefficients of the unique polynomial solution  $P^*(x)$  can be determined by solving the system (7.9.3).

We now direct our attention to the problem of fitting a polynomial of degree  $n - 1$  to a prescribed function in some best sense. Explicitly, let  $f(x)$  be bounded over the interval  $\alpha \leq x \leq \beta$  and let  $P$  denote a generic polynomial of degree at most  $n - 1$ . We seek to find the polynomial  $P^*(x)$  of degree  $n - 1$  for which

$$\min_P \max_{\alpha \leq x \leq \beta} |f(x) - P(x)|$$

is achieved.

Let  $X = (x_1, \dots, x_{n+1})$ ,  $\alpha \leq x_i \leq \beta$ , and  $y_i = f(x_i)$ . Then  $\rho(X)$  will denote the solution to the approximation problem for the points  $x_i, y_i$ . If at least two of the  $x_i$  are equal, the corresponding  $y_i$  are equal and we can find a polynomial passing through all the points; hence in this case  $\rho = 0$ .

► **THEOREM 7.9.2.** Let  $\rho_0 = \sup_X \rho(X)$ . Then

$$\min_P \max_{\alpha \leq x \leq \beta} |f(x) - P(x)| = \rho_0.$$

*Proof.* For any fixed  $P$  and  $X$ ,

$$\min_P \max_{\alpha \leq x \leq \beta} |f(x) - P(x)| \geq \min_P \max_i |f(x_i) - P(x_i)| = \rho(X).$$

Therefore

$$\min_P \max_{\alpha \leq x \leq \beta} |f(x) - P(x)| \geq \sup_X \rho(X) = \rho_0.$$

In order to prove the opposite inequality, for each  $x$  in  $[\alpha, \beta]$  we define the set

$$A_x = \left\{ \mathbf{a} = \langle a_0, \dots, a_{n-1} \rangle \left| \left| f(x) - \sum_{i=0}^{n-1} a_i x^i \right| \leq \rho_0 \right. \right\}.$$

Each  $A_x$  is a closed convex set, not necessarily bounded. By the definition of  $\rho_0$  we know that any  $n + 1$  of these sets intersect. If we take  $n + 1$  distinct  $x_i$ 's, the intersection of the corresponding sets is a parallelepiped, and therefore is bounded. It follows from Corollary 7.9.1 that they all intersect, and the theorem is established.

Suppose that in Theorem 7.9.2 we add the assumption that  $f(x)$  is continuous on  $\alpha \leq x \leq \beta$ . It is not hard to show that  $\rho(x_0, x_1, \dots, x_n)$  is continuous on  $\alpha \leq x_0 \leq x_1 \leq \dots \leq x_n \leq \beta$ . Thus we have the following corollary.

► **COROLLARY 7.9.2.** If  $f(x)$  is continuous on  $\alpha \leq x \leq \beta$ , there exist  $n + 1$  points  $(x_0^*, x_1^*, \dots, x_n^*)$  as described above such that

$$\min_{P_{n-1}} \max_{\alpha \leq x \leq \beta} |f(x) - P_{n-1}(x)| = \min_{P_{n-1}} \max_{x_\nu^*} |f(x_\nu^*) - P_{n-1}(x_\nu^*)|.$$

If we choose  $f(x) = x^n$  ( $\alpha = 0, \beta = 1$ ), then the solution of the problem is well known to be  $-x^n + T_{n-1}(x)$ , where  $T_{n-1}(x) = \cos((n-1)\theta)$ , with  $\cos \theta = x$ . In this case  $x_{\nu-1}^*$  turns out to be those  $x$  for which  $\theta = \pi/\nu$  ( $\nu = 1, \dots, n$ ) and  $x_n = 1$ .

Another application of Helly's Theorem is concerned with the covering of the unit sphere in  $E^n$  by closed hemispheres. A family of such hemispheres is called compact if the unit normals from the origin to the hyperplanes bounding the hemispheres, directed into the hemisphere, constitute a compact set.

► **THEOREM 7.9.3 (Covering Theorem).** Let the surface of a sphere in  $E^n$  be covered by a compact family of closed hemispheres; then we can find  $m$  ( $m \leq n + 1$ ) members of the family which cover the surface.

*Proof.* Let  $l_\alpha$  be the unit normal to the hemisphere  $H_\alpha$  with the orientation as specified. A point  $\mathbf{x}$  on the surface of the sphere is covered by  $H_\alpha$  if and only if  $(\mathbf{x}, l_\alpha) \geq 0$ .

Let  $\{l_i\}$  be a countable set dense in  $\{l_\alpha\}$ , and let  $\Gamma_m$  be the compact set spanned within the unit sphere by  $l_1, \dots, l_m$ . We wish to show that for  $m$  sufficiently large,  $\Gamma_m$  is arbitrarily close to the origin  $\theta$ . Otherwise for some  $\epsilon > 0$  the distance  $d(\theta, \Gamma_i)$  exceeds  $\epsilon$  for all  $i$ . By choice of  $\{l_i\}$ , the last statement implies that  $d(\theta, \Gamma) \geq \epsilon$ , where  $\Gamma$  is the convex set spanned by  $\{l_\alpha\}$ . Since  $\Gamma$  is convex, we can find a plane through the origin that does not pass within  $\epsilon$  of  $\Gamma$ ; if  $\mathbf{x}_0$  is the unit normal to this plane directed away from  $\Gamma$ , then  $(l_\alpha, \mathbf{x}_0) \leq -\epsilon$  for all  $\alpha$ . It follows that  $\mathbf{x}_0$  is not covered by  $H_\alpha$  for any  $\alpha$ , which contradicts our hypothesis.

This contradiction implies that for each  $k$  there exists  $m(k)$  such that  $d(\theta, \Gamma_{m(k)}) < 1/k$ . Let  $\mathbf{x}^k$  be a point of  $\Gamma_{m(k)}$  of distance less than  $1/k$  from the origin. By Lemma B.2.1, we have a convex representation

$$\mathbf{x}^k = \sum_{i=1}^{n+1} \xi_i^{(k)} l_i^{(k)}.$$

Since  $n + 1$  is fixed and  $\xi_i^{(k)}, l_i^{(k)}$  are drawn from compact sets, we may pass to the limit and obtain a representation

$$\theta = \sum_{i=1}^{n+1} \xi_i l_i,$$

with  $\xi_i \geq 0$ ,  $\sum \xi_i = 1$ . The hemispheres  $H_i$  corresponding to the  $l_i$  of the representation must cover the full sphere.

The theorem is not true if the compactness requirement is removed; for

example, the family of hemispheres in  $E^2$  described by the angles  $\pi, 1, \frac{1}{2}, \dots, 1/m, \dots$  covers the full circle but no finite subfamily will cover it.

The following corollary shows that one of the hemispheres may be chosen at will.

► **COROLLARY 7.9.3.** Let a given hemisphere  $H$  on the surface of a sphere in  $E^n$  be covered by a compact family of closed hemispheres. Then we can find  $m$  ( $m \leq n$ ) members of the family that cover  $H$ .

*Proof.* The given family, together with the closed complement  $H_0$  of  $H$ , covers the sphere. The last theorem guarantees that we can find a subfamily of the augmented family consisting of at most  $n + 1$  members which also covers the sphere. If this subfamily does not contain  $H_0$ , we consider the convex set  $C$  spanned within the unit sphere by the unit normals  $l_i$  of the subfamily.  $C$  must contain the origin  $\theta$ , since the subfamily covers the sphere. Let  $l_0$  denote the unit normal to  $H_0$ , and  $y_0$  the intersection of the radius  $[\theta, -l_0]$  with the boundary of  $C$ . Then  $y_0$  is a convex combination of  $n$  (or fewer) of the  $l_i$ , and  $\theta$  is a convex combination of  $y_0$  and  $l_0$ . (If  $y_0$  and  $\theta$  happen to coincide, then  $l_0$  will appear vacuously.) It follows that an  $(n + 1)$ -member subfamily containing  $l_0$  and covering the sphere can always be found. The closed complement of  $H_0$  must be covered by the other  $n$  members of the family.

Most of these statements about convex sets are essentially based on the proposition that any point belonging to the convex hull of a set  $H$  in  $E^n$  can be represented as a convex combination of at most  $n + 1$  points of  $H$ .

Another fundamental representation theorem which applies to interior points of convex sets is the following.

► **THEOREM 7.9.4.** Let  $C$  denote the convex hull of a finite set of points  $H$  in  $E^n$ . If  $z$  is a point of  $C$  such that there exists an open sphere about  $z$  in  $C$ , then there exists a subset  $H'$  of  $H$  consisting of at most  $2n$  points and such that its convex hull has  $z$  as an interior point.

*Proof.* The proof is by induction on the dimension of  $E^n$ . Suppose that the result is true for convex sets in Euclidean spaces of dimension less than  $n$ . The case of  $E^1$  is trivial. Choose  $m$  points  $y_i$  in  $H$  such that

$$z_0 = \sum_{i=1}^m \lambda_i y_i \quad (\lambda_i > 0; \sum \lambda_i = 1)$$

and  $2 \leq m \leq n + 1$ . Since  $z_0$  is interior, it is certain that  $m \geq 2$ ; the upper bound comes from the classical representation (Lemma B.2.1). The  $y_i$  may be selected so that the linear variety  $L$  determined by these points is  $(m - 1)$ -dimensional. Through  $z_0$  select any  $(n - m + 1)$ -dimensional linear manifold  $M$  independent of  $L$ . Project  $C$  perpendicularly in  $M$ . The

image  $\text{Proj}(H)$  of  $H$  in  $M$  spans convexly the projection of  $C$ . Within the space  $M$ , the point  $\mathbf{z}_0$  is also interior to  $\text{Proj}(C)$ . By the induction hypothesis there exist at most  $2(n - m + 1)$  points of  $\text{Proj}(H)$  whose convex span contains  $\mathbf{z}_0$  on its interior (in the sense of the space  $M$ ). To each such manifold  $M$  are associated at most  $2(n - m + 1)$  points of  $\text{Proj}(H)$  with the properties specified above. The points may be chosen as continuous functions of  $M$ ; hence, by a connectedness argument, we can obtain a single set of no more than  $2(n - m + 1)$  points  $\mathbf{z}_i$  which furnish the desired representation of  $\mathbf{z}_0$  for any manifold  $M$  independent of  $\mathbf{z}_0$ . These  $\mathbf{z}_i$ , which together with  $\mathbf{y}_i$  total at most  $2(n - m + 1) + m = 2n - m + 2 \leq 2n$  (since  $m \geq 2$ ), span a convex set  $C'$  with  $\mathbf{z}_0$  on its interior. Otherwise, there must be a hyperplane  $Y$  through  $\mathbf{z}_0$  that keeps  $C'$  on one side. Let  $\mathbf{l}$  be the perpendicular direction away from  $Y$  not including  $C'$ , and let  $M_0$  be any linear manifold of dimension  $2(n - m + 1)$  with  $M_0 \cap L = \mathbf{z}_0$  containing the direction  $\mathbf{l}$ . Then the projection of  $\mathbf{z}_i$  in  $M_0$  spans a convex set with  $\mathbf{z}_0$  noninterior, which contradicts the construction of  $\mathbf{z}_i$ .

A typical application of Theorem 7.9.4 is given in the following corollary.

► **COROLLARY 7.9.4.** Let a unit sphere  $S$  in  $E^n$  be covered by a finite family of open hemispheres. There exists a subfamily consisting of at most  $2n$  open hemispheres which also covers  $S$ .

The reader is referred to the problems at the close of this chapter for further applications.

We close this section with a constructive method of obtaining a finite representation. Let  $S_m$  denote the unit sphere of  $m$ -dimensional Euclidean space. We assume (1) that a closed set  $X = \{\mathbf{x}_\alpha\}$  of points on  $S_m$  is given, and (2) that the origin is interior to the convex set spanned by the points  $\mathbf{x}_\alpha$ . Assumption (2) is equivalent to the statement (2') that there exists a number  $d > 0$  such that, for every point  $\mathbf{y} \in S_m$ ,  $(\mathbf{x}_\alpha, \mathbf{y}) \geq d$  for some  $\alpha$ . What is desired is a way of choosing in a finite number of steps a finite subset  $\{\alpha_i\}$  satisfying (2) or (2').

*The method.* Assume (1) and (2') as stated above, and proceed according to the following recursive instructions:

- (0) Start with any point  $\mathbf{x}_1 \in X$ .
- (1<sub>k</sub>) Having chosen  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ , choose the point  $\mathbf{y}_k \in S_m$  which minimizes  $\max [(\mathbf{y}, \mathbf{x}_1), (\mathbf{y}, \mathbf{x}_2), \dots, (\mathbf{y}, \mathbf{x}_k)]$ .
- (2<sub>k</sub>) Choose  $\mathbf{x}_{k+1}$  as a point in  $X$  that maximizes  $(\mathbf{y}_k, \mathbf{x})$ .

The purpose of (1<sub>k</sub>) is to determine the point on the sphere farthest from those already chosen. So long as the origin is not in the interior of the convex set determined by  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ , this point will be uniquely defined, and will in fact be proportional to the (unique) nearest point of that set to the origin.

Step  $(2_k)$  selects a point of  $X$  nearest to  $\mathbf{y}_k$ . The way in which this is accomplished must depend on the way  $X$  is described. Since  $X$  is closed, the selection is always possible.

The process is stopped as soon as the minimum of  $(1_k)$  is positive. This will occur at some finite value of  $k$ .

*Proof that the method terminates.* Suppose that the method does not terminate, and define

$$m_k = \min_y \max [(\mathbf{y}, \mathbf{x}_1), (\mathbf{y}, \mathbf{x}_2), \dots, (\mathbf{y}, \mathbf{x}_k)],$$

$$n_k = (\mathbf{y}_k, \mathbf{x}_{k+1}).$$

Then, clearly,

$$-1 = m_1 \leq m_2 \leq \dots \leq m_k \leq \dots \leq 0$$

and  $n_k \geq d$  for all  $k$ . Making use of the compactness of  $S_m$  and  $X$ , select a subsequence  $k_1, k_2, \dots, k_l, \dots$  for which  $y_{k_l}$  and  $x_{k_l+1}$  both converge. Then

$$\lim_{l \rightarrow \infty} m_{k_l} \geq \lim_{l \rightarrow \infty} (\mathbf{y}_{k_l}, \mathbf{x}_{k_l+1}) = \lim_{l \rightarrow \infty} n_{k_l} \geq d > 0.$$

Hence  $m_{k_0}$  is positive for some finite  $k_0$ , contrary to the hypothesis that the method did not terminate.

### 7.10. Problems.

1. Consider the following nonlinear programming problem: Maximize  $g(\mathbf{x})$  with  $\mathbf{x} \in E^n$  satisfying the vector constraints  $\mathbf{F}(\mathbf{x}) \geq 0$  but otherwise unrestricted,  $g(\mathbf{x})$  concave, and  $\mathbf{F}(\mathbf{x})$  composed of concave functions. Assuming the condition of Theorem 7.1.1, show that the equivalent saddle-value problem has a kernel

$$\phi(\mathbf{x}, \mathbf{u}) = g(\mathbf{x}) + (\mathbf{u}, \mathbf{F}(\mathbf{x})),$$

where  $\mathbf{u} \geq \mathbf{0}$  and  $\mathbf{x}$  varies unrestrictedly.

2. Consider the problem

$$\max g(\mathbf{x}) \quad \text{subject to} \quad f(\mathbf{x}) \geq 0, \quad \mathbf{x} \geq \mathbf{0},$$

where  $f$  and  $g$  are concave functions. Prove that  $\mathbf{x}^0$  is optimal if and only if  $\mathbf{x}^0$  is a solution to

$$\max_x (\mathbf{g}_{x^0}, \mathbf{x}) \quad \text{subject to} \quad f(\mathbf{x}) \geq 0, \quad \mathbf{x} \geq \mathbf{0}$$

( $\mathbf{g}_{x^0}$  represents the vector of partial derivatives evaluated at  $\mathbf{x}^0$ ).

3. Consider the following problem: Maximize  $(\mathbf{p}, \mathbf{x})$  subject to  $\mathbf{x} \geq \mathbf{0}$ ,  $\mathbf{F}(\mathbf{x}) \leq \mathbf{v}$ , where  $F$  is a convex differentiable vector function and homogeneous of order one and  $\mathbf{p}$  is a fixed constant vector. Prove that if Theorem 7.1.1 is applicable, then  $(\mathbf{p}, \mathbf{x}^0) = (\mathbf{u}^0, \mathbf{v})$  for any saddle point  $\{\mathbf{x}^0, \mathbf{u}^0\}$ .

4. Prove that the linear programming problem

$$\max (\mathbf{c}, \mathbf{x}) \quad \text{subject to} \quad \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0},$$

where  $\mathbf{A}$  is a matrix, reduces to the quadratic programming problem

$$\max [(\mathbf{c}, \mathbf{x}) - \frac{1}{2}\epsilon(\mathbf{x}, \mathbf{x})] \quad \text{subject to} \quad \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

when  $\epsilon$  is a sufficiently small positive number.

5. Let  $f(t)$  be a monotone-increasing convex function with  $f(0) > 0$ . Find the set of  $x$ 's which minimize

$$\sum_{i=1}^n f(x_i)$$

subject to the restrictions

$$\sum_{i=1}^n x_i = T, \quad x_i \geq 0, \quad n \geq 1.$$

(Use Lagrange multipliers for each  $n$ .)

6. Solve the preceding problem with "convex" replaced by "concave."

7. Let  $f(x)$  be a continuous, strictly increasing convex function for nonnegative real  $x$ . Let  $0 \leq a_1 \leq a_2 \leq \dots \leq a_n$ , and define  $m$  as the largest index such that

$$\frac{a_m}{m} = \max_{1 \leq j \leq m} \frac{a_j}{j}.$$

Show that the solution of

$$\min_{\mathbf{x}} \sum_{k=1}^n f(x_k), \quad \text{where} \quad \sum_{j=1}^i x_j \geq a_i \quad \text{and} \quad x_i \geq 0,$$

has the property that  $x_1 = x_2 = \dots = x_m = a_m/m$ .

8. Determine inductively the complete solution of Problem 7.

9. Consider the following problem: Maximize  $(\mathbf{a}, \mathbf{x}) - \frac{1}{2}(\mathbf{x}, \mathbf{Ax})$  subject to  $\mathbf{Bx} \leq \mathbf{b}$ , where  $\mathbf{A}$  is a positive-definite matrix and  $\mathbf{B}$  is an arbitrary matrix. Prove that a solution to this problem, if one exists, can be obtained by solving the problem

$$\max [(\mathbf{c}, \mathbf{y}) - \frac{1}{2}(\mathbf{y}, \mathbf{Dy})] \quad \text{subject to} \quad \mathbf{y} \geq \mathbf{0},$$

where  $\mathbf{c} = \mathbf{DA}^{-1}\mathbf{a} - \mathbf{b}$  and  $\mathbf{D} = \mathbf{BA}^{-1}\mathbf{B}'$ .

10. Let  $h(v) = \max_x \{g(\mathbf{x}) | f(\mathbf{x}) \leq v, \mathbf{x} \geq \mathbf{0}\}$ , where  $f$  is convex and  $g$  is concave. Prove that  $h(v)$  is concave and that for any shadow price vector  $\langle u_1^0, \dots, u_m^0 \rangle$

$$\left( \frac{\partial h}{\partial v_k} \right)_+ \leq u_k^0 \leq \left( \frac{\partial h}{\partial v_k} \right)_- \quad (k = 1, \dots, m).$$

(Assume that Theorem 7.1.1 can be applied.)

11. Let  $\mathbf{G}(\mathbf{x})$  and  $\mathbf{F}(\mathbf{x})$  denote two families of vector functions all of whose components consist of concave functions defined for all  $\mathbf{x}$  in  $E^n$ . Consider the problem: Find  $\mathbf{x}^0$  satisfying

$$\max_{\mathbf{x}} \min_i (\mathbf{G}(\mathbf{x}))_i,$$

where  $\mathbf{F}(\mathbf{x}) \geq \mathbf{0}$ ,  $\mathbf{x} \geq \mathbf{0}$ . We call this problem the minimum-component maximum problem. Prove that  $\mathbf{x}^0$  cannot be a solution of the minimum-component maximum problem unless there exists a nonnegative vector  $\mathbf{v}^0$  ( $\sum v_i^0 = 1$ ) satisfying  $(\mathbf{v}^0, \mathbf{G}(\mathbf{x}^0)) = \min_i [\mathbf{G}(\mathbf{x}^0)]_i$  and such that  $\{\mathbf{x}^0, \mathbf{u}^0\}$ , where  $\mathbf{u}^0 > \mathbf{0}$ , is a saddle point of

$$\phi(\mathbf{x}, \mathbf{u}) = (\mathbf{v}^0, \mathbf{G}(\mathbf{x})) + (\mathbf{u}, \mathbf{F}(\mathbf{x})).$$

12. In the statement of Theorem 7.1.1 the following condition was imposed: For each  $\mathbf{u} > \mathbf{0}$  there exists  $\mathbf{x}$  in  $X$  satisfying  $(\mathbf{u}, \mathbf{F}(\mathbf{x})) > 0$ . Prove that this is equivalent to the existence of an element  $\mathbf{x}^0$  in  $X$  for which  $\mathbf{F}(\mathbf{x}^0) \gg \mathbf{0}$ .

13. Consider the problem

$$\max_{\mathbf{x}} [(\mathbf{c}, \mathbf{x}) - \frac{1}{2}(\mathbf{x}, \mathbf{Ax})] \quad \text{subject to} \quad \mathbf{x} \geq \mathbf{0},$$

where  $\mathbf{A}$  is positive-definite. Prove that the iterative relation

$$x_i^{(u+1)} = \max \left\{ 0, \frac{1}{a_{ii}} \left( c_i - \sum_{j < i} a_{ij} x_j^{(u+1)} - \sum_{j \geq i} a_{ij} x_j^{(u)} \right) \right\} \quad (i = 1, \dots, n; u = 0, 1, 2, \dots),$$

where  $\langle x_1^0, x_2^0, \dots, x_n^0 \rangle \geq \mathbf{0}$ , converges to a solution.

14. Consider the problem

$$\max_{\mathbf{x}} g(\mathbf{x}) \quad \text{subject to} \quad \mathbf{F}(\mathbf{x}) \geq \mathbf{0}, \quad \mathbf{x} \geq \mathbf{0},$$

where  $\mathbf{F}$  is a vector function. Let  $K = \{k | f_k(\mathbf{x}) = 0 \text{ for all feasible } \mathbf{x}\}$ . Assume (1) that  $f_k(\mathbf{x})$  is linear in  $\mathbf{x}$  for  $k \in K$ ; (2) that  $df_k/dx$  is linearly independent for  $k \in K$ ; (3) that for any  $i$ , there is a feasible vector  $\mathbf{x}^i$  such that  $x_i^i > 0$ . Prove that the original problem reduces to the problem of finding a saddle point for  $g(\mathbf{x}) + (\mathbf{u}, \mathbf{F}(\mathbf{x}))$  in  $\mathbf{x} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}$ .

The next three problems employ the notation of Sections 7.5 and 7.6.

15. In Theorem 7.6.1, when is  $V = [\phi_1, \Gamma_1] \cup [\phi_2, \Gamma_2]$  closed? Show that the set  $V$  will be closed whenever  $C_1$  and  $C_2$  have common relative interior points.
16. Show that  $\sup f_\alpha(\mathbf{x})$  exists for some  $\mathbf{x}$  if and only if  $[f_\alpha, C_\alpha]$  have a common point or  $[\phi_\alpha, \Gamma_\alpha]$  have a common nonvertical barrier.
17. Prove that whenever the directional derivatives of  $\partial f/\partial x$  and  $\partial g/\partial x$  are uniformly bounded, there exists a nonvertical hyperplane separating  $[g, D]$  and  $[f + \mu, C]$  of Theorem 7.8.1, even if  $C$  and  $D$  have no common relative interior points.
18. Deduce the min-max theorem of matrices (Theorem 1.4.1) from Theorem 7.8.1.
19. Using the theory of conjugate functions, prove that

$$\xi x \leq \frac{|x|^p}{p} + \frac{|\xi|^q}{q},$$

where  $\xi$  and  $x$  are real numbers and

$$1/p + 1/q = 1 (p > 1).$$

20. Let a unit sphere  $S$  in  $E^n$  be covered by a finite family of open hemispheres. Show that at most  $2n$  open hemispheres of this family cover  $S$  (this is Corollary 7.9.4).
21. Let  $H_\alpha$  denote a finite family of open half-spaces with boundary through the origin which covers all of  $E^n$ . Prove that at most  $2n$  members of  $H_\alpha$  cover all of  $E^n$ .
22. Let  $M_\alpha$  denote a finite family of closed half-spaces in  $E^n$  with the origin on the boundary of each  $M_\alpha$  such that every  $2n$  members of  $M_\alpha$  have a common ray. Prove that all members of  $M_\alpha$  possess a common ray.
23. A finite family  $R_\alpha$  of rays in  $E^n$  is said to be irreducible if for any hyperplane through the origin there is at least one ray properly on each side of the hyperplane. Prove that if  $R_\alpha$  is a finite irreducible family of rays in  $E^n$ , then there exists a subfamily of  $R_\alpha$  consisting of at most  $2n$  rays which is also irreducible.
24. Let  $S$  and  $S'$  be two finite collections of points in  $E^n$  such that for every  $n + 2$  points  $R$  of  $S \cup S'$  there exists a hyperplane strictly separating  $S' \cap R$  from  $S \cap R$ . Prove that there exists a hyperplane which strictly separates  $S$  from  $S'$ .
25. Let  $S$  and  $S'$  be two finite collections of points in  $E^n$  such that for any  $2n + 2$  points  $T$  of  $S \cup S'$ , there exists a hyperplane which separates  $S \cap T$  from  $S' \cap T$ . Show that there exists a hyperplane separating  $S$  from  $S'$ .

## NOTES AND REFERENCES TO CHAPTER 7

**§ 7.1.** The first major development in the theory of nonlinear programming (N.L.P.) was the fundamental paper by Kuhn and Tucker [164]. Their work may more appropriately be called concave programming, which is the terminology accepted in this book.

The Kuhn-Tucker analogs of Theorems 7.1.1 and 7.1.2 depend on the differentiability properties of the functions, whereas our methods depend only on the separation properties of disjoint convex sets. Consequently, the hypotheses of Theorem 7.1.1 are not equivalent or comparable to those of Kuhn and Tucker. Theorem 7.1.3 is taken directly from their paper.

Other modifications of the Kuhn-Tucker theorems have been presented by Arrow and Hurwicz [9].

Two collections of contributions to the theory of N.L.P. and linear inequalities are [163] and [12], the first edited by Kuhn and Tucker and the second edited by Arrow, Hurwicz, and Uzawa.

**§ 7.2.** The mathematical problem of portfolio selection was discussed by Markowitz [172]. We partly follow his treatment. Markowitz is interested in the more general problem of determining efficient vector portfolio investment schemes in which the first component is the expected yield and the second component is the variance of the portfolio selection.

The blending of octane numbers of gasoline mixtures is a natural example of concave programming (see Manne [171]).

The theory of consumer buying abounds with examples of N.L.P. (see Samuelson [211]).

A N.L.P. version of foreign trade and exchange has been studied by Reiter [204].

The interindustry activity analysis model discussed in this section was formulated by Chenery [47] (see also Chenery and Kretschmer [48]). The iterative solution was proposed by Chenery and Uzawa [49].

Many important nonlinear programming models come from production and inventory studies in which the cost of production is a convex function of the amount produced. The original investigation of this type of problem was made by Hohn and Modigliani [125]. An extensive study of these models is carried out in Arrow, Karlin, and Scarf [16, Chapters 4-7].

**§ 7.3.** The gradient method of equations (7.3.1) was proposed by Arrow and Hurwicz [10], who proved the convergence to a pair of saddle points in the small. This means that if the starting pair of feasible vectors is sufficiently close to a saddle point, the solution of the gradient equations converges. Uzawa [235] established the convergence for any initial position. We have followed his proof. The gradient method of this section may be construed as a natural extension of the Brown-Von Neumann differential-equations scheme of Section 6.7.

With regard to computing procedures, we call attention to the simplex method for quadratic programming developed by Wolfe [256]. Alternative computational methods that are sometimes valid for quadratic programming problems have been proposed by Markowitz [173] and Houthakker [127].

**§ 7.4.** The vector maximum problem and the corresponding theorems are due originally to Kuhn and Tucker [164], who rely on differentiability arguments in contradistinction to the methods used here.

In connection with the vector maximum problem, the reader is referred to the work of Hurwicz [128] on programming in abstract spaces (see also Kretschmer [154]).

**§ 7.5.** The interesting theory of conjugate functions was discovered by Fenchel. An exposition of this theory is contained in his book *Convex Cones, Sets, and Functions* [81]. Fenchel develops the duality of conjugate functions from the algebraic point of view, using the notion of polar reciprocity. Our presentation of this subject is strictly geometric.

**§ 7.6.** Theorem 7.6.4 is due to Bohnenblust, Karlin, and Shapley [37]. The proof here is new and uses the ideas of conjugate functions.

**§ 7.8.** The results of this section are all due to Fenchel [81].

**§ 7.9.** Theorems on fitting points by polynomials of fixed degree have a long history. Theorem 7.9.2 is due to Rademacher and Schoenberg [202]. Most of the other results of this section can be found in Karlin and Shapley [140].

**§ 7.10.** In connection with Problems 5–8, see Charnes and Cooper [45] and Savage [214]. Problem 9 is due to Hildreth [122]. Problem 11 is based on results first obtained by Kuhn and Tucker [164]. The result of Problem 12 was noted by Uzawa. Problem 13 is due to Hildreth [122]. Problem 14 was suggested by Uzawa [12].

## CHAPTER 8

### MATHEMATICAL METHODS IN THE STUDY OF ECONOMIC MODELS

In the past two decades, impressive progress has been made in the mathematical analysis of economic models. It has been necessary to develop new methods to deal with such models, since they typically involve nondifferentiable functions and variables subject to inequality constraints, which cannot be handled by the classical calculus. Broadly speaking, the methods needed are those used in game theory—i.e., the theory of convex sets, topological fixed-point theorems, the theory of positive matrices, etc. In this chapter and the next, we explore the use of these techniques in the context of mathematical economics.

In this chapter we shall examine elements of production theory, consumption theory, and equilibrium theory. In Chapter 9 we shall consider welfare economics, models of the dynamic theory of balanced growth, and certain problems of stability theory associated with equilibrium prices. We do not formally consider macro-economic models, aggregation theory, and monopoly-oligopoly studies, since the mathematical aspects of these subjects are not too well developed (see Problems 1–6).

In short, we have not attempted to present an exhaustive study of mathematical economics. Rather, we have striven to give as complete a description as possible of the intrinsic mathematical methodology of economic models, and to point out the relevance of these techniques to the broader concept of decision processes.

In Section 8.1 we formulate various simple linear models of equilibrium and exchange. We arrive at solutions of these linear models by appealing to the theory of positive matrices, which is discussed at length in Section 8.2. Positive matrices, aside from occurring naturally in economic studies, appear also as part of the theory of finite stochastic processes in the investigation of some mechanical systems; hence the results of Section 8.2 are of fundamental importance. Also in this section, which may be read independently of the rest of the chapter, we describe some properties of matrices whose eigenvalues have negative real parts. These properties (see pp. 247–56) are relevant to our discussion of stability theory in Chapter 9.

Prior to a general investigation of the competitive equilibrium, it seems necessary to elaborate separately on the two main components of equilibrium theory, namely production theory and consumption theory. In Section 8.4 we describe the nature of the mathematical problems connected

with production theory, and in Section 8.5 we examine the problem of characterizing efficient points for activity analysis models, i.e., the problem of selecting an optimal production process from among several alternative processes.

In Section 8.6 we discuss the theory of consumption. Our approach is to postulate a preference relation, from which we derive a demand function. Since this approach is independent of the specific form of the utility index function of the consumer, we can formulate Slutsky relations which describe the variations of the demand function solely in terms of price changes and changes in the level of the budget constraint for a given preference relation. The concept of revealed preference is also discussed in this section.

In the final sections of this chapter, production and consumption theory are combined in the analysis of general static equilibrium models. We treat several different versions of equilibrium, culminating in the Arrow-Debreu model. The mathematically oriented reader will be impressed by the depth of the topological techniques that are used in analyzing the equations of supply and demand.

**8.1 Open and closed linear Leontief models.** One of the simplest models of exchange for an economy can be described as follows:

There are  $n$  activities (or industries), each producing a product. Let  $X_i$  represent the total output of activity  $i$ , and let  $x_{ij}$  represent the total amount of the product of activity  $i$  used by the  $j$ th activity. If

$$Y_i = X_i - \sum_{j=1}^n x_{ij},$$

then  $Y_i$  expresses the difference between the total output of activity  $i$  and the amount of its product consumed by the  $n$  activities.  $Y_i$  is commonly referred to as the “final demand” of commodity  $i$ .

In a *closed Leontief model*  $Y_i = 0$  for all  $i$ . There are no surpluses. If labor is a type of activity, the total labor force is employed; if steel is another type of activity, all of the steel produced is consumed by the system. A closed model typically includes such activities as services, management, and the production of raw materials.

In an *open Leontief model*  $Y_i > 0$  for at least one  $i$ . From an economic viewpoint an open model implies the presence of exogenous material—some quantity of capital, labor, raw material, etc., that is supplied to the system from the outside. The essential characteristic of the open model is the existence of outside demand or supply or both.

We define

$$a_{ij} = \frac{x_{ij}}{X_j},$$

which may be interpreted as the amount of activity  $i$ 's commodity needed to produce one unit of commodity  $j$ . The quantities  $a_{ij}$  are referred to as the "production coefficients," and are assumed in this section to be constants independent of the actual values of  $X_i$ . (This assumption is of doubtful validity and should be thought of only as a first approximation.) It follows that

$$(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{Y}, \quad (8.1.1)$$

where  $\mathbf{A} = \|a_{ij}\|$ ,  $\mathbf{X} = \langle X_1, X_2, \dots, X_n \rangle$ , and  $\mathbf{Y} = \langle Y_1, Y_2, \dots, Y_n \rangle$ .

In practice the matrix  $\mathbf{A}$  is assumed to be known or estimable. This is a severe assumption over which economists have fired off many a polemic.

The most frequent problem concerning the open Leontief model is as follows: Given that the  $Y_i$  are known, find  $X_i$  satisfying (8.1.1). Of course, the answer is meaningful only if the resulting solution  $\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$  exists and is a nonnegative vector.

Let  $(\mathbf{I} - \mathbf{A})^{-1} = \|A_{ij}\|$ . Then  $X_i = \sum_j A_{ij}Y_j$  ( $i = 1, 2, \dots, n$ ), so that  $A_{ij}$  can be interpreted as the amount by which the output of activity  $i$  must be increased to produce one additional unit of commodity  $j$ .

Another problem involving the open model is the following: Consider  $n$  countries  $1, 2, \dots, n$ . Let  $a_{ij}$  denote country  $j$ 's marginal propensity to import from country  $i$  (that is, the increase in imports from country  $i$  to country  $j$  per unit increase in income of country  $j$ ), and let  $a_{ii}$  denote country  $i$ 's marginal propensity to consume its own goods. Let  $X_i$  represent country  $i$ 's national income, and let  $c_i$  represent country  $i$ 's national expenditure, which is independent of income. Then

$$X_i = \sum_{j=1}^n a_{ij}X_j + c_i,$$

and hence  $(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{c}$ . To be economically meaningful,  $\mathbf{X}$  and  $\mathbf{c}$  must be positive vectors.

This example may also be interpreted as a closed model of trade between countries. Suppose that all income to country  $j$  comes from the sale of goods to other countries or to itself. Let  $\alpha_{ij}$  represent the fraction of country  $j$ 's income that is spent on goods from country  $i$ . Let  $X_j$  be the annual income of country  $j$ . The total value of exports from country  $i$  is clearly

$$\sum_{j=1}^n \alpha_{ij}X_j,$$

and this, by the definition of a closed system, is  $X_i$ . We thus have

$$\sum_{j=1}^n \alpha_{ij}X_j = X_i \quad (i = 1, \dots, n),$$

where

$$\sum_{i=1}^n \alpha_{ij} = 1, \quad \alpha_{ij} \geq 0.$$

One problem to be solved in dealing with equations of this sort is to determine conditions which assure a positive solution.

In the case of a closed Leontief model, we seek a positive solution  $\mathbf{X}$  of the system of equations  $\mathbf{AX} = \mathbf{X}$ . Nontrivial solutions exist if and only if the value 1 is an eigenvalue of the matrix  $\mathbf{A}$ . But even when 1 is an eigenvalue, to ensure the existence of a positive eigenvector requires something more. The theory of positive matrices serves to resolve such problems. Several conditions for the solvability of these problems are developed in the next section and applied in Section 8.3.

The usual dual problems in an economic model involving labor and prices make their appearance in the system of equations generated by the transpose matrix of  $\mathbf{A}$ . Suppose that the  $(n + 1)$ th activity represents labor. If there is no unemployment, then

$$X_{n+1} - \sum_{j=1}^n X_{n+1,j} = 0.$$

As before, we define

$$a_{n+1,j} = \frac{X_{n+1,j}}{X_j},$$

so that  $a_{n+1,j}$  can be interpreted as the amount of labor used to produce a single unit of commodity  $j$ . It follows that  $\sum_i a_{n+1,i} A_{ij}$  is the additional amount of labor needed to produce one more unit of commodity  $j$ .

Let  $p_i$  be the price of a unit of commodity  $i$ . What is usually assumed is that the price of a commodity is equal to the cost of producing it, i.e., that

$$p_j = \sum_{i=1}^N p_i a_{ij} + p_{n+1} a_{n+1,j} \quad \text{or} \quad \mathbf{p}(\mathbf{I} - \mathbf{A}) = p_{n+1}(\mathbf{a}_{n+1}).$$

Here  $\mathbf{a}_{n+1}$  represents the vector  $\langle a_{n+1,1}, a_{n+1,2}, \dots, a_{n+1,n} \rangle$ . This is a sort of equilibrium condition; it may be thought of as resulting from competition so complete as to make profit impossible. Ordinarily  $p_{n+1}$  is taken to be 1, so that the  $p_i$  represent relative prices.

**†8.2 The theory of positive matrices.** The study of matrices with positive elements is very important for a variety of applications. Such matrices often arise in a natural way—i.e., the nature of the model implies the posi-

---

† The results of this section will only be used in Sections 8.3 of this chapter and in Sections 9.3, 9.7, and 9.8 of Chapter 9.

tivity of the coefficients—in probability models (particularly the study of Markov chains), economic analyses, the study of vibrations of mechanical systems, etc. We accordingly digress in this section to display in separate form several propositions concerning eigenvalues of positive matrices that will be pertinent to our analysis of economic models.

Suppose that  $\mathbf{A} = \|a_{ij}\|$  is a square matrix. If all  $a_{ij}$  are nonnegative, we write  $\mathbf{A} \geq \mathbf{0}$ ; if  $\mathbf{A} \geq \mathbf{0}$  and some  $a_{ij} > 0$ , we write  $\mathbf{A} > \mathbf{0}$ ; and if all  $a_{ij} > 0$ , we write  $\mathbf{A} \gg \mathbf{0}$ . This is consistent with the notation and terminology adopted earlier for vectors.

We write  $\mathbf{A} \geq \mathbf{B}$  if  $\mathbf{A} - \mathbf{B} \geq \mathbf{0}$ , etc. A positive matrix as a mapping clearly preserves the partial ordering defined by  $>$ . In particular, if  $\mathbf{x}' \geq \mathbf{x}$ , then  $\mathbf{x}'\mathbf{A} \geq \mathbf{x}\mathbf{A}$  for  $\mathbf{A} \geq \mathbf{0}$ . Let

$$\Lambda = \{\lambda | \lambda\mathbf{A} \geq \lambda\mathbf{x} \quad \text{for some } \mathbf{x} > \mathbf{0}\}. \quad (8.2.1)$$

The set always contains the value  $\lambda = 0$  whenever  $\mathbf{A} \geq \mathbf{0}$ . The relation  $\lambda\mathbf{A} \geq \lambda\mathbf{x}$  being homogeneous, we may without loss of generality assume that the nonnegative vector  $\mathbf{x}$  corresponding to a value  $\lambda$  is normalized so that  $\sum x_i = 1$ . Such vectors  $\mathbf{x}$  conveniently traverse a compact set. Let  $\lambda_0 = \sup_{\Lambda} \lambda$ . We obviously have  $\lambda_0 < \infty$ . We begin by an analysis of matrices with strictly positive elements and then pass to the modifications that are necessary in dealing with general positive matrices.

►THEOREM 8.2.1. If  $\mathbf{A} \gg \mathbf{0}$ , then (a) there exists  $\mathbf{x}^0 \gg \mathbf{0}$  such that  $\mathbf{x}^0\mathbf{A} = \lambda_0\mathbf{x}^0$ ; (b) if  $\lambda \neq \lambda_0$  is any other eigenvalue of  $\mathbf{A}$ , it follows that  $|\lambda| < \lambda_0$ ; (c)  $\lambda_0$  is an eigenvalue of geometrical and algebraic multiplicity 1.

*Remark 8.2.1.* To clarify the meaning of the theorem, we make the following assertions, which are proved in Section A.4. If

$$N_{\lambda}^r = [\mathbf{x} | \mathbf{x}(\mathbf{A} - \lambda\mathbf{I})^r = \mathbf{0}] \quad \text{and} \quad M_{\lambda}^r = [\mathbf{x} = \mathbf{y} | \mathbf{y}(\mathbf{A} - \lambda\mathbf{I})^r | \mathbf{y} \quad \text{arbitrary}],$$

then  $N_{\lambda}^0 \subset N_{\lambda}^1 \subset N_{\lambda}^2 \subset \cdots \subset N_{\lambda}^k = N_{\lambda}^{k+1} = N_{\lambda}^{k+2} = \cdots$  for any  $\mathbf{A}$  and  $\lambda$ , where  $\subset$  means “is contained in but not equal to.” The integer  $k$  thus determined is called the index of  $\lambda$ . The dimension of  $N_{\lambda}^1$  is the geometric multiplicity of  $\lambda$ , and the dimension of  $N_{\lambda}^k$  is the algebraic multiplicity of  $\lambda$ . Also  $M_{\lambda}^0 \supset M_{\lambda}^1 \supset M_{\lambda}^2 \supset \cdots \supset M_{\lambda}^k = M_{\lambda}^{k+1} = M_{\lambda}^{k+2} = \cdots$  for the same  $k$ . For  $\mathbf{A}$  self-adjoint,  $k = 1$  for any eigenvalue  $\lambda$ , and hence its geometric and algebraic multiplicities are the same. In any case the geometric multiplicity is no larger than the algebraic multiplicity. When the algebraic and geometric multiplicities agree, the elementary divisors are simple.

*Proof of the Theorem.*

(a) A simple compactness argument implies the existence of a vector  $\mathbf{x}^0 > \mathbf{0}$  such that  $\mathbf{x}^0\mathbf{A} \geq \lambda_0\mathbf{x}^0$ . Since  $\mathbf{A} \gg \mathbf{0}$  implies that  $\mathbf{x}\mathbf{A} \gg \mathbf{0}$  for

$\mathbf{x} > \mathbf{0}$ , it follows that  $\mathbf{x}^0 \mathbf{A} = \mathbf{y}^0 \gg \mathbf{0}$  satisfies  $\mathbf{y}^0 \mathbf{A} \gg \lambda_0 \mathbf{y}^0$  unless  $\mathbf{x}^0 \mathbf{A} = \lambda_0 \mathbf{x}^0$ . But if  $\mathbf{y}^0 \mathbf{A} \gg \lambda_0 \mathbf{y}^0$ , a positive  $\epsilon$  can be found such that  $\mathbf{y}^0 \mathbf{A} \geq (\lambda_0 + \epsilon) \mathbf{y}^0$ , thus contradicting the meaning of  $\lambda_0$ . Hence  $\mathbf{x}^0 \mathbf{A} = \lambda_0 \mathbf{x}^0$  and  $\mathbf{x}^0 \gg \mathbf{0}$ .

(b) Let  $\lambda \neq \lambda_0$  be an eigenvalue of  $\mathbf{A}$  and let the corresponding non-trivial eigenvector  $\mathbf{z}$  be such that  $\mathbf{z} \mathbf{A} = \lambda \mathbf{z}$ . Let  $|\mathbf{z}|$  denote the vector whose components are  $\langle |z_i| \rangle$ . Then the positivity of  $\mathbf{A}$  implies that  $|\mathbf{z}| \mathbf{A} \geq |\lambda| |\mathbf{z}|$ . From the construction of  $\lambda_0$ , it follows that  $|\lambda| \leq \lambda_0$ . In order to see that strict inequality persists, we consider the matrix  $\mathbf{A}_\delta = \mathbf{A} - \delta \mathbf{I}$ , where  $\delta$  is chosen so small and positive that  $\mathbf{A}_\delta$  is a strictly positive matrix. The largest real positive eigenvalue of  $\mathbf{A}_\delta$  is  $\lambda_0 - \delta$ , which as we have seen bounds (in magnitude) all other eigenvalues of  $\mathbf{A}_\delta$ . If the eigenvalue satisfies  $|\lambda| = |\lambda_0|$ , with  $\lambda$  distinct from  $\lambda_0$ , then  $|\lambda - \delta| > |\lambda_0 - \delta|$ , contradicting the previous conclusion. The proof of the theorem is complete.

(c) Suppose that the set of all eigenvectors with eigenvalue  $\lambda_0$  is at least two-dimensional. Since  $\mathbf{A}$  is a real matrix, there exists a real eigenvector  $\mathbf{z}$  that is linearly independent of  $\mathbf{x}^0$ . Since  $\mathbf{x}^0 \gg \mathbf{0}$ , a linear combination  $\mathbf{w} = t \mathbf{x}^0 + \mathbf{z}^0$  exists ( $t = \min_i - z_i^0 / x_i^0$ ), with  $\mathbf{w} > \mathbf{0}$  but not  $\mathbf{w} \gg \mathbf{0}$ . Since  $\mathbf{w} \mathbf{A} = \lambda_0 \mathbf{w}$  is strictly positive and cannot be identically zero, we have an absurdity. Hence we conclude that the geometrical multiplicity of  $\lambda_0$  is 1.

Suppose that  $\mathbf{y}(\mathbf{A} - \lambda_0 \mathbf{I}) = \mathbf{z}$  is such that  $\mathbf{z}(\mathbf{A} - \lambda_0 \mathbf{I}) = \mathbf{0}$ . By the first part of this argument  $\mathbf{z}$  is a multiple of  $\mathbf{x}^0$ . On multiplying  $\mathbf{y}$  by a nonzero constant, we may assume that  $\mathbf{z} = a \mathbf{x}^0$ , with  $a \geq 0$ . Let  $\mathbf{f}^0$  denote a positive eigenvector for the matrix  $\mathbf{A}'$  corresponding to the eigenvalue  $\lambda_0$ ; we may verify the existence of  $\mathbf{f}^0$  by applying assertion (a) of the theorem to the strictly positive matrix  $\mathbf{A}'$ . Observe that

$$(a \mathbf{x}^0, \mathbf{f}^0) = (\mathbf{y}(\mathbf{A} - \lambda_0 \mathbf{I}), \mathbf{f}^0) = (\mathbf{y}, (\mathbf{A}' - \lambda_0 \mathbf{I}) \mathbf{f}^0) = 0,$$

so that since  $(\mathbf{x}^0, \mathbf{f}^0) > 0$  we must have  $a = 0$ . Hence if  $\mathbf{y}(\mathbf{A} - \lambda_0 \mathbf{I})^2 = \mathbf{0}$ , then  $\mathbf{y}(\mathbf{A} - \lambda_0 \mathbf{I}) = \mathbf{0}$ , and thus  $N_{\lambda_0}^2 = N_{\lambda_0}^1$ . This means that the algebraic multiplicity of  $\lambda_0$  is equal to the geometric multiplicity of  $\lambda_0$ .

In order to explore in greater detail the structure of the eigenmanifold attached to the eigenvalue  $\lambda_0$ , we introduce a projection operation which projects vectors into the linear space of eigenvectors associated with  $\lambda_0$ . For this purpose, let the eigenvector  $\mathbf{x}^0$ , where  $\mathbf{x}^0 \mathbf{A} = \lambda_0 \mathbf{x}^0$ , and the eigenvector  $\mathbf{f}^0 \gg \mathbf{0}$ , where  $\mathbf{A} \mathbf{f}^0 = \lambda_0 \mathbf{f}^0$ , be normalized so that  $\sum x_i^0 f_i^0 = 1$ . Define  $\mathbf{P}$  as the matrix  $\|x_i^0 f_i^0\|$ . Then  $\mathbf{P}$  has the following properties:

- (i) For any vectors  $\mathbf{x}$  and  $\mathbf{f}$ ,  $\mathbf{x} \mathbf{P} = (\mathbf{x}, \mathbf{f}^0) \mathbf{x}^0$  and  $\mathbf{P} \mathbf{f} = (\mathbf{f}, \mathbf{x}^0) \mathbf{f}^0$ .
- (ii)  $\mathbf{P}^2 = \mathbf{P}$ .
- (iii)  $\mathbf{A} \mathbf{P} = \mathbf{P} \mathbf{A} = \lambda_0 \mathbf{P}$ .

The verification of these assertions is immediate.

For any matrix  $\mathbf{B}$ , it is known that  $(\mathbf{B} - \lambda \mathbf{I})^{-1}$  represents an analytic function of  $\lambda$  (i.e., each component is an analytic function of  $\lambda$ ) wherever the inverse makes sense. Each eigenvalue of  $\mathbf{B}$  gives rise to a pole of  $(\mathbf{B} - \lambda \mathbf{I})^{-1}$  (i.e., a pole for some component) whose order is related to the algebraic multiplicity of the eigenvalue. Specifically, the order of the pole at an eigenvalue  $\tilde{\lambda}$  is equal to the index of  $\tilde{\lambda}$  (see Remark 8.2.1 above and pp. 383–85 of Appendix A). Moreover,

$$(\mathbf{B} - \lambda \mathbf{I})^{-1} = - \sum_{n=0}^{\infty} \frac{1}{\lambda^{n+1}} \mathbf{B}^n$$

converges for  $|\lambda| > r$ , where  $r$  denotes the radius of the smallest circle containing the eigenvalues of  $\mathbf{B}$ , also called the spectral radius. (We frequently write  $\lambda_0(\mathbf{A}) = r(\mathbf{A})$  to denote the spectral radius of  $\mathbf{A}$ .) But  $(\mathbf{B} - \lambda \mathbf{I})^{-1}$  has a pole at some value  $\lambda^*$ , where  $|\lambda^*| = \lambda_0$ ; hence

$$r = \lim_{n \rightarrow \infty} \sqrt[n]{\max_{ij} |b_{ij}^{(n)}|},$$

where  $\mathbf{B}^n = \|b_{ij}^{(n)}\|$ . We are now prepared to prove the following theorem.

► **THEOREM 8.2.2.** For  $\mathbf{A} \gg \mathbf{0}$ ,  $\lambda_0$ , and  $\mathbf{P}$  as defined above,

$$\frac{1}{\lambda_0^n} \mathbf{A}^n \rightarrow \mathbf{P}.$$

*Proof.* The eigenvalues of the matrix  $\mathbf{B} = \mathbf{A} - \lambda_0 \mathbf{P}$  must all lie strictly within the circle of radius  $\lambda_0$ , since (i) any nonzero eigenvalue of  $\mathbf{B}$  is an eigenvalue of  $\mathbf{A}$ , and (ii)  $\lambda_0$  is not an eigenvalue of  $\mathbf{B}$ .

We now prove these statements:

Let us assume that these two assertions are false. Then  $\mathbf{zB} = \lambda \mathbf{z}$  for some nonzero  $\mathbf{z}$ . It follows that  $\mathbf{zBP} = \lambda \mathbf{zP}$  and

$$\mathbf{zBP} = \mathbf{z}(\mathbf{AP} - \lambda_0 \mathbf{P}^2) = \lambda_0 \mathbf{z}(\mathbf{P} - \mathbf{P}) = \mathbf{0}.$$

Therefore, if  $\lambda \neq 0$ , it follows that  $\mathbf{zP} = \mathbf{0}$  and  $\mathbf{zB} = \lambda \mathbf{z}$  reduces to  $\mathbf{zA} = \lambda \mathbf{z}$ . In particular, if  $\mathbf{zB} = \lambda_0 \mathbf{z}$ , then  $\mathbf{zP} = \mathbf{0}$  and  $\mathbf{zA} = \lambda_0 \mathbf{z}$ . But then  $\mathbf{z}$  is a nonzero multiple of  $\mathbf{x}^0$  and  $\mathbf{zP}$  cannot be zero. This contradiction shows that the assertions are true. The spectral radius of  $\mathbf{B}$  must be smaller than  $\lambda_0$ , and it follows that the maximum component of  $\mathbf{B}^m$  satisfies the bound  $|b_{ij}^{(m)}| < K\rho^m$  for all  $m$  with a suitable  $\rho < \lambda_0$ . (See the remarks preceding this theorem.) But  $\mathbf{B}^m = \mathbf{A}^m - \lambda_0^m \mathbf{P}$ ; hence we infer that  $\mathbf{A}^m / \lambda_0^m$  converges to  $\mathbf{P}$  at the geometric convergence rate  $\rho / \lambda_0$ .

All of the foregoing theory holds if  $\mathbf{A} \geq \mathbf{0}$  and  $\mathbf{A}^m \gg \mathbf{0}$  for some  $m$ . It also holds if  $\mathbf{A}$  maps some convex cone in  $E^n$  into its interior. The specific cone associated with positive matrices is the first quadrant.

Several of the above results hold whenever  $\mathbf{A} \geq \mathbf{0}$ , i.e., even when we do not insist that some iterate of  $\mathbf{A}$  be strictly positive. Let  $\mathbf{U}$  be the square matrix with every component 1; then  $\mathbf{A} + \delta\mathbf{U}$  ( $\delta > 0$ ) is strictly positive and the assertions of Theorem 8.2.1 apply. A standard limiting argument on  $\delta$  will establish the validity of parts (a) and (b) of the following theorem.

►THEOREM 8.2.3. Let  $\mathbf{A} > \mathbf{0}$  and let  $\lambda_0$  be defined as in (8.2.1). Then (a)  $\lambda_0$  is an eigenvalue and there exists  $\mathbf{x}^0 > \mathbf{0}$  such that  $\mathbf{x}^0\mathbf{A} = \lambda_0\mathbf{x}^0$ ; (b) if  $\lambda$  is any other eigenvalue,  $|\lambda| \leq \lambda_0$ ; (c)

$$\frac{1}{m} \sum_{i=1}^m \frac{\mathbf{A}^i}{\lambda_0^i}$$

converges whenever  $\mathbf{x}^0 \gg \mathbf{0}$ ; and (d) if  $\lambda$  is an eigenvalue of  $\mathbf{A}$  with  $|\lambda| = \lambda_0$ , it follows that  $\eta = \lambda/\lambda_0$  is a root of unity and  $\eta^m\lambda_0$  is an eigenvalue of  $\mathbf{A}$  for  $m = 0, 1, 2, \dots$

*Proof of (c).* For convenience of exposition we assume that  $\lambda_0 = 1$ . We first note that the elements of  $\mathbf{A}^m$  are uniformly bounded. In fact,  $\mathbf{x}^0\mathbf{A}^m = \mathbf{x}^0$  implies

$$0 \leq a_{ij}^{(m)} \leq \frac{\max_i x_i^0}{\min_i x_i^0},$$

as desired.

Let  $\mathcal{L}$  denote the space of elements for which  $\mathbf{z}\mathbf{A} = \mathbf{z}$ . Clearly,  $\mathcal{L}$  is a closed linear space such that for  $\mathbf{z}$  in  $\mathcal{L}$

$$\mathbf{z} \frac{\mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^m}{m} = \mathbf{z}$$

converges as  $m$  increases to infinity. Next define  $\mathcal{K} = \{\mathbf{y} | \mathbf{y} = \mathbf{z}(\mathbf{I} - \mathbf{A})\}$ ; that is,  $\mathcal{K}$  is the range of  $\mathbf{I} - \mathbf{A}$ . We assert that for  $\mathbf{y}$  in  $\mathcal{K}$ ,  $\mathbf{y}\mathbf{S}_m$  tends to zero, where

$$\mathbf{S}_m = \frac{\mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^m}{m}.$$

Explicitly, since  $\mathbf{y} = \mathbf{z}(\mathbf{I} - \mathbf{A})$ , it follows that  $\mathbf{y}\mathbf{S}_m = (\mathbf{z}\mathbf{A} - \mathbf{z}\mathbf{A}^{m+1})/m$  tends to zero, by virtue of the fact that the coefficients of  $\mathbf{A}^{m+1}$  are uniformly bounded.

We have thus established that  $\mathbf{S}_m$  converges for all vectors contained in the sum  $\mathcal{L} + \mathcal{K}$ . It remains only to show that  $\mathcal{L} + \mathcal{K}$  is all of the Euclidean  $n$ -space. To this end, we observe that every vector  $\mathbf{x}$  can be written as

$$\mathbf{x} = (\mathbf{x} - \mathbf{x}\mathbf{S}_m) + \mathbf{x}\mathbf{S}_m = \mathbf{y}_m + \mathbf{z}_m.$$

Moreover, as  $m$  tends to infinity we may select a limit vector  $\mathbf{z}_m$  of  $\mathbf{z}_m$ .

This is possible since  $S_m$  consists of a sequence of uniformly bounded matrices. It is immediately verified that  $\mathbf{z}_0$  is in  $\mathcal{L}$ . Furthermore, it is readily verified that each  $\mathbf{y}_m$  belongs to  $\mathcal{K}$  and hence its limit belongs to  $\mathcal{K}$ . This finishes the proof of (c).

*Proof of (d).* We may assume without loss of generality that there exists some  $\mathbf{f}^0 \gg \mathbf{0}$  for which  $\mathbf{A}\mathbf{f}^0 = \mathbf{f}^0$ . The reason for this is as follows: Suppose for definiteness that  $\mathbf{f}^0$  has the form  $\langle f_1^0, f_2^0, \dots, f_r^0, 0, \dots, 0 \rangle$ , where  $f_1^0 > 0, \dots, f_r^0 > 0$  [ $\mathbf{f}^0$  exists by (a) above applied to the matrix  $A'$ ]. Since  $\mathbf{A}\mathbf{f}^0 = \mathbf{f}^0$ , it follows that

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{B} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix},$$

where  $\mathbf{A}_1$  is a positive  $r \times r$  matrix and  $\mathbf{A}_2$  is a positive  $(n-r) \times (n-r)$  matrix. To investigate the nature of the eigenvalues of  $\mathbf{A}$  it is clearly sufficient to study the eigenvalues of  $\mathbf{A}_1$  and  $\mathbf{A}_2$  separately. By continuing in this way, we ultimately arrive at the problem for which the assumption made concerning  $\mathbf{f}^0$  is valid.

We now proceed to the proof of (d). Suppose that  $\lambda$  is an eigenvalue of  $\mathbf{A}$  of modulus 1 and that  $\mathbf{x}\mathbf{A} = \lambda\mathbf{x}$  for some nonzero vector  $\mathbf{x}$ , that is,

$$\sum_{i=1}^n x_i a_{ij} = \lambda x_j.$$

Then  $\sum_i |x_i| a_{ij} \geq |x_j|$  or  $|\mathbf{x}|\mathbf{A} \geq |\mathbf{x}|$ . But

$$(\mathbf{f}^0, |\mathbf{x}|) \leq (\mathbf{f}^0, |\mathbf{x}|\mathbf{A}) = (\mathbf{A}\mathbf{f}^0, |\mathbf{x}|) = (\mathbf{f}^0, |\mathbf{x}|),$$

with strict inequality yielding a contradiction unless  $|\mathbf{x}|\mathbf{A} = |\mathbf{x}|$ . Consequently,

$$\sum_{i=1}^n |x_i| a_{ij} = \left| \sum_{i=1}^n x_i a_{ij} \right|.$$

This implies the existence of numbers  $u_j$  ( $|u_j| = 1$ ) such that

$$x_i a_{ij} = |x_i| a_{ij} u_j \quad (*)$$

for each  $i$  and  $j$ . Let  $\mathbf{x} \cdot \mathbf{y}$  denote the vector whose components are  $\{x_i y_i\}$ . Multiplying the preceding relation by  $u_i$  and summing shows that

$$(\mathbf{x} \cdot \mathbf{u})\mathbf{A} = (\mathbf{u} \cdot |\mathbf{x}|)\mathbf{A} \cdot \mathbf{u}$$

and from (\*) by summation on  $i$

$$|\mathbf{x}| \cdot \mathbf{u} = (|\mathbf{x}|\mathbf{A}) \cdot \mathbf{u} = \mathbf{x}\mathbf{A} = \lambda\mathbf{x}.$$

Consider the vector  $\mathbf{x} \cdot \mathbf{u}$ . Using the last two relations proves that

$$(\mathbf{x} \cdot \mathbf{u})\mathbf{A} = (\mathbf{u} \cdot |\mathbf{x}|)\mathbf{A} \cdot \mathbf{u} = \lambda \mathbf{x}\mathbf{A} \cdot \mathbf{u} = \lambda^2 \mathbf{x} \cdot \mathbf{u},$$

and hence  $\mathbf{u} \cdot \mathbf{x}$  is an eigenvector corresponding to  $\lambda^2$ . Furthermore,

$$u_i^2 |x_i| a_{ij} u_j = u_i^2 x_i a_{ij},$$

and hence  $(|\mathbf{x}| \cdot \mathbf{u}^2)\mathbf{A} \cdot \mathbf{u} = (\mathbf{x} \cdot \mathbf{u}^2)\mathbf{A}$ . But

$$(|\mathbf{x}| \cdot \mathbf{u}^2)\mathbf{A} = \lambda(\mathbf{x} \cdot \mathbf{u})\mathbf{A} = \lambda^3 \mathbf{u} \cdot \mathbf{x},$$

and thus  $(\mathbf{x} \cdot \mathbf{u}^2)\mathbf{A} = \lambda^3(\mathbf{x} \cdot \mathbf{u}^2)$ . This shows that  $\mathbf{x} \cdot \mathbf{u}^2$  is an eigenvector for  $\lambda^3$ . Proceeding in this way, we find that if  $\lambda$  is an eigenvalue of modulus 1, then  $\lambda$  is a root of unity and  $\lambda, \lambda^2, \lambda^3, \dots, \lambda^n, \dots$  are all eigenvalues as asserted. An eigenvector of  $\lambda^{r+1}$  is  $\mathbf{x} \cdot \mathbf{u}^r$ , where  $\mathbf{x}$  and  $\mathbf{u}$  have the meaning indicated above. This completes the proof of Theorem 8.2.3.

The following characterization of the spectral radius  $\lambda_0(\mathbf{A})$ , which is implicit in the construction of  $\Lambda$  in (8.2.1), was used in the proofs of Theorems 8.2.1 and 8.2.3. It is useful enough in general to be worth highlighting here.

► **COROLLARY 8.2.1.** The eigenvalue of  $\mathbf{A}$  of largest magnitude  $\lambda_0 = \lambda_0(\mathbf{A})$  is real and nonnegative and is characterized as  $\lambda_0 = \max_{\Lambda} \lambda$ , where

$$\Lambda = \{\lambda | \mathbf{x}\mathbf{A} \geq \lambda \mathbf{x} \quad \text{for some } \mathbf{x} > \mathbf{0}\}. \quad (8.2.1)$$

Another useful consequence is as follows:

► **COROLLARY 8.2.2.** If there exists  $\mathbf{x}^0 \gg \mathbf{0}$  such that  $\mathbf{x}^0 \mathbf{A} \leq \mu \mathbf{x}^0$ , then  $\mu$  is an upper bound to the spectral radius of  $\mathbf{A}$ .

*Proof.* It quickly follows that

$$|a_{ij}^{(n)}| \leq \mu^n \frac{\max_i x_i^0}{\min_i x_i^0},$$

where  $\mathbf{A}^n = \|a_{ij}^n\|$ ; hence

$$\overline{\lim} \sqrt[n]{|a_{ij}^{(n)}|} \leq \mu,$$

which leads easily to the desired result (see the discussion preceding Theorem 8.2.2).

The next corollary follows from Corollary 8.2.1.

► **COROLLARY 8.2.3.** If  $\mathbf{A} \geq \mathbf{B} \geq \mathbf{0}$  and  $\lambda_0(\mathbf{A}), \lambda_0(\mathbf{B})$  are defined as in (8.2.1), corresponding to  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, then  $\lambda_0(\mathbf{A}) \geq \lambda_0(\mathbf{B})$ ; i.e.,

the spectral radius of a positive matrix is a monotone-increasing function of its components.

The following corollary may be compared with the fundamental min-max theorem of game theory.

►COROLLARY 8.2.4. If  $\mathbf{A} \geq 0$  and  $\mu_0 = \inf \{\mu | \mathbf{Af} \leq \mu \mathbf{f}, \mathbf{f} > \mathbf{0}\}$ , and if the eigenvector  $\mathbf{x}^0$  is strictly positive, then

$$\mu_0 = \lambda_0 = \sup_{\substack{x > 0 \\ \sum x_i = 1}} \inf_{\substack{f > 0 \\ \sum f_i = 1}} \frac{(\mathbf{f}, \mathbf{x}\mathbf{A})}{(\mathbf{f}, \mathbf{x})} = \inf_{\substack{f > 0 \\ \sum f_i = 1}} \sup_{\substack{x > 0 \\ \sum x_i = 1}} \frac{(\mathbf{f}, \mathbf{x}\mathbf{A})}{(\mathbf{f}, \mathbf{x})},$$

where it is understood that if both  $(\mathbf{f}, \mathbf{x}) = 0$  and  $(\mathbf{f}, \mathbf{x}\mathbf{A}) = 0$  hold, the meaning of  $(\mathbf{f}, \mathbf{x}\mathbf{A})/(\mathbf{f}, \mathbf{x})$  is taken to be  $\lambda_0$ .

*Proof.* Since there exists a positive eigenvector  $\mathbf{f}^0$  (Theorem 8.2.3) corresponding to  $\mathbf{A}'$  for the eigenvalue  $\lambda_0$ , it follows that  $\mu_0 \leq \lambda_0$ .

But for  $\mathbf{f}$  corresponding to  $\mu_0$ , which certainly exists, we obtain

$$\mu_0(\mathbf{f}, \mathbf{x}^0) \geq (\mathbf{Af}, \mathbf{x}^0) = (\mathbf{f}, \mathbf{x}^0\mathbf{A}) = \lambda_0(\mathbf{f}, \mathbf{x}^0).$$

The hypothesis on  $\mathbf{x}^0$  compels  $\mu_0 \geq \lambda_0$ . Hence  $\mu_0 = \lambda_0$ . The remaining relations follow in the usual way (see Theorem 1.3.1).

The following theorem is useful for its applications to economics.

►THEOREM 8.2.4. If  $\rho > \lambda_0(\mathbf{A})$  and  $\mathbf{A} \geq 0$ , then  $(\rho \mathbf{I} - \mathbf{A})^{-1}$  transforms nonnegative vectors into nonnegative vectors.

*Proof.* For  $\lambda > \lambda_0(\mathbf{A})$ , where  $\lambda_0(\mathbf{A})$  is the spectral radius, an explicit calculation yields the representation

$$(\lambda \mathbf{I} - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{\lambda^{k+1}}.$$

Since each  $\mathbf{A}^k$  maps the nonnegative orthant into itself, so does the infinite sum.

The converse of Theorem 8.2.4 is also valid: i.e., if  $(\rho \mathbf{I} - \mathbf{A})^{-1}$  maps the nonnegative orthant into itself, then  $\rho$  is an upper bound to the spectral radius of  $\mathbf{A}$ . Explicitly, if  $\mathbf{x}\mathbf{A} \geq \lambda \mathbf{x}$  for  $\mathbf{x} > \mathbf{0}$ , then  $\mathbf{z} = \mathbf{x}(\mathbf{A} - \rho \mathbf{I}) + (\rho - \lambda)\mathbf{x}$  is a nonnegative vector. Hence  $(\rho \mathbf{I} - \mathbf{A})^{-1}$  applied to  $\mathbf{z}$  yields a nonnegative vector, or

$$-\mathbf{x} + (\rho - \lambda)\mathbf{x}(\mathbf{A} - \rho \mathbf{I})^{-1} \geq \mathbf{0}.$$

But  $\mathbf{x} > \mathbf{0}$  and  $\mathbf{x}(\mathbf{A} - \rho \mathbf{I})^{-1} \geq \mathbf{0}$ , which implies that  $\rho > \lambda$ . On the other hand,  $\lambda_0$  is such an admissible  $\lambda$  value (Corollary 8.2.1), and hence  $\rho > \lambda_0(\mathbf{A})$  as asserted.

We may express the conditions under which  $(\rho\mathbf{I} - \mathbf{A})^{-1}$  transforms positive vectors into positive vectors in terms of restrictions on the principal minors of  $(\rho\mathbf{I} - \mathbf{A})$  as follows: If every principal minor of  $\rho\mathbf{I} - \mathbf{A}$  is positive, then  $(\rho\mathbf{I} - \mathbf{A})^{-1}$  is positivity-preserving.

To prove this statement it is enough by virtue of Theorem 8.2.4 to show that  $\rho > \lambda_0(\mathbf{A})$ . This is done as follows: The relation  $\rho = \lambda_0(\mathbf{A})$  is impossible, since by hypothesis  $\det(\rho\mathbf{I} - \mathbf{A}) > 0$ . In view of Theorem 8.2.4, it is enough to show that  $\det(t\mathbf{I} - \mathbf{A}) \neq 0$  for  $t > \rho$ . Clearly,

$$\det(t\mathbf{I} - \mathbf{A}) = \det[(t - \rho)\mathbf{I} + \mathbf{B}(\rho)], \quad (8.2.2)$$

where  $\mathbf{B}(\rho) = \rho\mathbf{I} - \mathbf{A}$ . Expanding the second terms of (8.2.2), we see because of the hypothesis that  $\det(t\mathbf{I} - \mathbf{A}) > 0$  for  $t > \rho$  as desired.

The hypothesis can be weakened so that the assertion reads as follows: If a single sequence of  $n$  “bordered” principal minors  $|\rho\mathbf{I} - \mathbf{A}|_1^p$  ( $p = 1, \dots, n$ ) is positive, then every principal minor of  $\rho\mathbf{I} - \mathbf{A}$  is positive, and hence, by what has been established above,  $\rho$  exceeds the greatest characteristic number of the matrix  $\mathbf{A}$ . The proof of this assertion is omitted (see Problem 11).

*Matrix decomposability.* Another set of ideas connected with positive matrices involves the notion of matrix indecomposability. We consider  $\mathbf{A} = \|a_{ij}\|$ , a nonnegative  $n \times n$  matrix. A set  $S$  of indices will be said to be closed if  $a_{pq} = 0$  for  $q \in S, p \notin S$ . If no proper closed set exists, the matrix  $\mathbf{A}$  is said to be indecomposable. This will happen if and only if there exists no permutation matrix  $\mathbf{\Pi}^\dagger$  such that

$$\mathbf{\Pi} \mathbf{A} \mathbf{\Pi}^{-1} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_3 \\ 0 & \mathbf{A}_2 \end{bmatrix},$$

where  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are square submatrices.

A matrix is decomposable, as indicated in the proof of Theorem 8.2.3, part (d), whenever it has a nonnegative eigenvector, not strictly positive, of a positive eigenvalue. For instance, if  $\mathbf{A}\mathbf{y}^0 = \lambda_0\mathbf{y}^0$ , where  $\mathbf{y}^0 = \langle y_1^0, \dots, y_n^0 \rangle$ , is arranged so that  $y_1^0 > 0, \dots, y_r^0 > 0$  while  $y_{r+1}^0 = y_{r+2}^0 = \dots = y_n^0 = 0$ , then necessarily

$$\mathbf{A} = \frac{r}{n-r} \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix},$$

with  $\mathbf{A}_3 = \mathbf{0}$ .

---

† A permutation matrix is a matrix such that each row and each column has a single nonzero entry of value 1.

The class of positive indecomposable matrices may further be divided into an aperiodic subclass and a periodic subclass. It is elementary to show that an indecomposable matrix  $\mathbf{A}$  by permutation of rows and columns can be put in the form

$$\mathbf{\Pi} \mathbf{A} \mathbf{\Pi}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}_\omega \\ \mathbf{G}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G}_{\omega-1} & \mathbf{0} \end{bmatrix}, \quad (8.2.3)$$

where the  $\mathbf{0}$ 's on the diagonal are square matrices. The number  $\omega$  is related to the number of eigenvalues on the circle of radius  $\lambda_0$  [see Theorem 8.2.3, part (d)]. If  $\omega > 1$ , the indecomposable matrix  $\mathbf{A}$  is said to be periodic; if  $\omega = 1$ ,  $\mathbf{A}$  is said to be aperiodic.

Applications of some of these ideas to economic models will be found in the next section. Indecomposability is also an important concept related to the principle of communication between states of a Markov chain.

*Metzler matrices.* We close this section with a discussion of the eigenvalue structure of matrices  $\mathbf{C} = \|\mathbf{c}_{ij}\|$ , where  $\mathbf{c}_{ij} \geq 0$  ( $i \neq j$ ). Such matrices are referred to by economists as Metzler matrices; we shall call them  $M$ -matrices.  $M$ -matrices may be treated as a special class of positive matrices, since by adding a sufficiently large positive multiple of the identity matrix we can achieve

$$\mathbf{A} = \mathbf{C} + \mu \mathbf{I}$$

such that  $\mathbf{A} \geq \mathbf{0}$ .

We now direct our analysis to the problem of determining criteria for  $M$ -matrices which ensure that all their eigenvalues have negative real parts.  $M$ -matrices of this form are of great importance in economic studies, notably in the study of stability characteristics associated with price systems converging to equilibrium, since local stability is guaranteed for a matrix system of differential equations whenever all eigenvalues have negative real parts (see Section 9.3).

Since we only translate  $\mathbf{C}$ , all the eigenvalues of  $\mathbf{C}$  have negative real parts if and only if  $\Re(\lambda_i) < \mu$ , where  $\lambda_i$  are the eigenvalues of  $\mathbf{A}$ . Because of Theorem 8.2.3, since  $\mathbf{A} \geq \mathbf{0}$ , this will be so if and only if  $\lambda_0(\mathbf{A}) < \mu$ . Analogous arguments relying heavily on the theory of positive matrices allow us to deduce more, namely:

(a) The eigenvalue of an  $M$ -matrix with largest real part is real and has an associated nonnegative eigenvector. We define  $\sigma(\mathbf{C}) = \max_i \Re(\alpha_i)$ , where  $\alpha_i$  are the eigenvalues of  $\mathbf{C}$ . (This use of  $\sigma$  will be maintained hereafter.)

- (b)  $\sigma(\mathbf{C}) < 0$  if and only if there exists  $\mathbf{x} > \mathbf{0}$  such that  $\mathbf{Cx} \ll \mathbf{0}$ .  
 (c)  $\sigma(\mathbf{C}) < 0$  if and only if  $-\mathbf{C}^{-1}$  is positivity-preserving.

Assertions (a) and (b) follow immediately from Theorem 8.2.3. Assertion (c) follows from Theorem 8.2.4.

Assertion (c) can be expressed equivalently in terms of the minors of  $\mathbf{C}$  (see the parallel argument on p. 254). Explicitly,  $\sigma(\mathbf{C}) < 0$  if and only if

$$c_{11} < 0, \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0, \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} < 0, \dots, (-1)^n \begin{vmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{vmatrix} > 0. \quad (8.2.4)$$

The reader is referred to Problems 10–15 for further propositions concerning positive matrices and  $M$ -matrices and their applications.

**8.3 Applications of the theory of positive matrices to the study of linear models of equilibrium and exchange.** (1) In our discussion of closed Leontief models in Section 8.1, we did not prove the existence of positive solutions to the system of equations  $\mathbf{AX} = \mathbf{X}$ , where  $\mathbf{A}$  represents the matrix of production coefficients. (Recall that  $a_{ij} = x_{ij}/X_j$ , where  $x_{ij}$  is the total amount of the  $i$ th product used by the  $j$ th activity.) To ensure the existence of positive solutions naturally requires some additional conditions. One reasonable condition suggested by economic considerations is that  $\sum_i x_{ij} = X_j$ , i.e., that the total value output of the  $j$ th activity is fully used in acquiring the various inputs used. This may be thought of as being part of the definition of a closed economic system: for example, in an economy consisting solely of farmers, weavers, and carpenters, our assumption says that the farmer spends all his income on clothing, housing, and food.

The problem is to determine the amounts to be produced consistent with the production coefficients. In terms of the coefficients, our assumption leads to the relation

$$\sum_{i=1}^n a_{ij} = 1$$

for every  $j$ .

As a consequence, the matrix  $\mathbf{A}$  must have spectral radius 1. For if  $\mathbf{u} = \langle 1, 1, 1, \dots, 1 \rangle$ , with  $n$  components, then  $\mathbf{uA} = \mathbf{u}$ ; and if  $\mathbf{Az} = \lambda z$ , with  $\sum |z_i| = 1$ , then  $(\mathbf{u}, \mathbf{A}|z|) \geq |\lambda| \sum |z_i|$ . (Again  $|z|$  represents the vector with components  $|z_i|$ .) But  $(\mathbf{uA}|z|) = \sum |z_i|$ , from which it follows that  $|\lambda| \leq 1$ , and since the eigenvalues of  $\mathbf{A}$  and  $\mathbf{A}'$  coincide,  $\lambda_0(\mathbf{A}) = 1$ . It follows by Theorem 8.2.3 that there exists a nonzero, nonnegative vector  $\mathbf{X}$  which solves  $\mathbf{AX} = \mathbf{X}$ .

We have proved the following theorem.

►THEOREM 8.3.1. In a closed Leontief system with

$$\sum_{i=1}^n x_{ij} = X_j,$$

there exists a nontrivial positive solution to the equations  $\mathbf{AX} = \mathbf{X}$ . If  $\mathbf{A}$  has a strictly positive iterate, then the solution is unique except for a multiplying factor and possesses only strictly positive components.

The last part of the theorem is inferred from the statement of Theorem 8.2.1.

In the case of an open system, our problem is to determine conditions such that  $(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{Y}$  can be solved for any prescribed final demand vector  $\mathbf{Y}$  such that the solution  $\mathbf{X}$  is a nonnegative vector.

►THEOREM 8.3.2. If there exists some output vector  $\mathbf{x}^0$  such that  $(\mathbf{I} - \mathbf{A})\mathbf{x}^0$  is a strictly positive vector, then  $(\mathbf{I} - \mathbf{A})^{-1}$  exists and transforms nonnegative vectors into nonnegative vectors. (The expression  $(\mathbf{I} - \mathbf{A})^{-1}$  may be viewed as a linear transformation acting on either column or row vectors, whichever is appropriate. The positivity-preserving character of a matrix and that of its transpose are obviously equivalent.)

*Remark 8.3.1.* The hypothesis of this theorem means that there is some combination of outputs which guarantees a surplus for every commodity.

*Proof.* In view of Theorem 8.2.4 it is sufficient to show that all eigenvalues of  $\mathbf{A}$  are of magnitude smaller than 1. Let  $\lambda$  be such that  $\mathbf{zA} = \lambda\mathbf{z}$ , with  $\mathbf{z}$  a nonzero vector. The positivity of  $\mathbf{A}$  implies that  $|\mathbf{z}|\mathbf{A} \geq |\lambda||\mathbf{z}|$ . The hypothesis requires that

$$0 < ((\mathbf{I} - \mathbf{A})\mathbf{x}_0, |\mathbf{z}|) = (\mathbf{x}_0, |\mathbf{z}|) - (\mathbf{x}_0, |\mathbf{z}|\mathbf{A}),$$

and since  $\mathbf{x}_0$  is a nonnegative vector, it follows that  $0 < (\mathbf{x}_0, |\mathbf{z}|)(1 - |\lambda|)$ . Hence  $|\lambda| < 1$  as asserted.

The conclusion of Theorem 8.3.2 holds for any hypothesis strong enough to drive all the eigenvalues of  $\mathbf{A}$  into the interior of the unit circle of the complex plane. The hypothesis of Theorem 8.3.2 accomplishes this.

Another theorem of this kind, whose proof is a paraphrase of the previous argument, is the following.

►THEOREM 8.3.3. If there exists some price vector  $\mathbf{p}^0$  such that  $\mathbf{p}^0(\mathbf{I} - \mathbf{A})$  is strictly positive, then  $(\mathbf{I} - \mathbf{A})^{-1}$  transforms nonnegative vectors into nonnegative vectors.

(2) As an example of a situation in economics in which the concepts of indecomposability are meaningful, consider the model introduced on p. 245

involving export-import relations between countries. If the import matrix  $\mathbf{A}$  is indecomposable, then a dollar spent in any one country will eventually induce spending in every other country. In particular, if  $\mathbf{A}$  has the form of (8.2.3), then countries in  $G_i$  will spend only in those in  $G_{i+1}$ ; and after  $\omega$  instances of spending, the original expenditure by  $G_i$  in  $G_{i+1}$  will have an influence in  $G_i$ , giving rise to a sort of spending cycle.

(3) We close this section with a short description of a dynamic situation whose analysis depends on Theorem 8.2.2.

We may think of the model of exchange introduced on p. 245 as happening over time. In the original analysis the equations  $\mathbf{Ax} = \mathbf{x}$  describe a closed model in which the components  $x_i$  of  $\mathbf{x}$  represent the total output (or activity level) of the  $i$ th activity with no reference to a time period over which this equilibrium is maintained. These equations may be expanded and given a dynamic interpretation as follows. Let  $\mathbf{x}^1$  represent the initial vector of activity levels. Then  $\mathbf{Ax}^1 = \mathbf{x}^2$  can be regarded as the outputs attributable to the activities of one time period. Similarly,  $\mathbf{Ax}^2 = \mathbf{x}^3$  expresses the vector of the activity level required in the third time period, and  $\mathbf{Ax}^n = \mathbf{x}^{n+1}$  has the analogous interpretation. Since  $\mathbf{x}^{n+1} = \mathbf{A}^n \mathbf{x}^1$ , we find that under suitable assumptions, namely where Theorem 8.2.2 is applicable,  $\mathbf{x}^{n+1}$  converges to a limiting output vector  $\mathbf{x}^0$  such that  $\mathbf{x}^0 = \mathbf{Ax}^0$ . It is easy to impose criteria which guarantee that  $\mathbf{x}^0$  is a nonzero positive eigenvector of  $\mathbf{A}$  and possesses suitable meaning in the context of the closed model. Other convergence and stability results of this sort will be developed in Sections 9.2 through 9.6.

**8.4 The theory of production.** In this section we generalize the Leontief assumptions to describe a nonlinear model involving production.

The theory of production is concerned first with the allocation of productive factors among various technological activities to produce goods for consumption, and then with the distribution of the value of the total product among the productive factors.

Consider a situation in which  $n$  finished products are produced by using a given amount of each of  $r$  factors of production, where labor, land, raw material, etc. are considered factors of production. (We shall later discuss a production problem in which we make no differentiation between finished products and factors of production.) The products will be denoted by  $i = 1, 2, \dots, n$ , and the factors of production by  $k = 1, 2, \dots, r$ .

The technological possibility of production will be explained in terms of *activity analysis*. The economy has a finite set of basic activities  $j = 1, \dots, m$ , each of which has an “activity level” associated with it in any given set of circumstances. The activity level for the economy as a whole may be denoted by an  $m$ -vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_m \rangle$ , where each com-

ponent  $x_j$  is nonnegative and stands for the level at which the basic activity  $j$  is operated. The technology of the system will now be characterized by specifying a pair of vector functions  $\mathbf{f}(\mathbf{x})$  and  $\mathbf{g}(\mathbf{x})$ :  $\mathbf{f}(\mathbf{x}) = \langle f_1(\mathbf{x}), \dots, f_n(\mathbf{x}) \rangle$  is an  $n$ -vector, the  $i$ th component of which stands for the amount of product  $i$  produced by operating the system at activity level  $\mathbf{x}$ , while  $\mathbf{g}(\mathbf{x}) = \langle g_1(\mathbf{x}), \dots, g_r(\mathbf{x}) \rangle$  is an  $r$ -vector, the  $k$ th component of which stands for the amount of productive factor  $k$  required in order to operate at activity level  $\mathbf{x}$ .

Let us now assume that certain amounts of the productive factors, which are represented by an  $r$ -vector  $\mathbf{v} = \langle v_1, v_2, \dots, v_r \rangle$ , are available to the system, and that the system is concerned only with getting the maximum value of the output as evaluated at given market prices

$$(\mathbf{p} = \langle p_1, p_2, \dots, p_n \rangle, p_i \geq 0).$$

The problem of production can now be stated as follows: Find an activity level  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  that maximizes the value of output

$$(\mathbf{p}, \mathbf{f}(\mathbf{x})) = \sum_i p_i f_i(\mathbf{x}) \quad (8.4.1)$$

subject to the restrictions

$$\mathbf{x} \geq \mathbf{0}, \quad (8.4.2)$$

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{v}. \quad (8.4.3)$$

The allocation of productive factors among various activities thus formulated is a special version of one of the linear (or nonlinear) programming problems (see Section 7.1).

The basic assumptions for the technology, which play an important role in the economic analysis of production and will be assumed in what follows, are the laws of constant returns to scale and of the diminishing marginal rates of transformation. The former is expressed by the positive homogeneity of  $\mathbf{f}(\mathbf{x})$  and  $\mathbf{g}(\mathbf{x})$ , the latter by the concavity of  $\mathbf{f}(\mathbf{x})$  and the convexity of  $\mathbf{g}(\mathbf{x})$ .

Invoking the theory of nonlinear programming (Section 7.1), we may reduce the problem of production to the problem of finding a saddle point  $\{\bar{\mathbf{x}}, \bar{\mathbf{\omega}}\}$  of the Lagrangian form

$$\phi(\mathbf{x}, \mathbf{\omega}) = (\mathbf{p}, \mathbf{f}(\mathbf{x})) + (\mathbf{\omega}, \mathbf{v} - \mathbf{g}(\mathbf{x})) \quad (\mathbf{x} \geq \mathbf{0}, \mathbf{\omega} \geq \mathbf{0}).$$

The vector  $\bar{\mathbf{\omega}}$  of the solution of the saddle-value problem may be interpreted as an imputed price vector; it can be construed as a relative measure of the distribution of the value of the total product among the productive factors (see p. 117).

A special case of the production problem occurs when there is no joint product, i.e., when the amount of each product produced is uniquely determined by the amounts of the productive factors used in the activity that produces it. By taking an appropriate scaling of the basic activities, we may state this special problem as follows: Find the distribution of the productive factors  $\langle v_{ik} \rangle$  that maximizes

$$\sum_{i=1}^n p_i f_i(v_{i1}, \dots, v_{ir})$$

subject to the restraints

$$v_{ik} \geq 0 \quad (i = 1, \dots, n; k = 1, \dots, r),$$

$$\sum_{i=1}^n v_{ik} \leq v_k \quad (k = 1, \dots, r),$$

where  $v_{ik}$  denotes the amount of the productive factor  $k$  used in activity  $i$ .

A typical example is the Cobb-Douglas model:

$$f_i(v_{i1}, \dots, v_{ir}) = \alpha_i v_{i1}^{\beta_{i1}}, \dots, v_{ir}^{\beta_{ir}},$$

where  $\alpha_i, \beta_{ik}$  are constants and  $\beta_{ik} \geq 0, \sum_k \beta_{ik} = 1$ .

*Production possibility sets.* The distinction between finished products (consumption goods) and factors of production is rather arbitrary. A finished product may be used in the production of other finished products, as in the classical Leontief model, and also a productive factor may be produced.

It is therefore of some purpose to formulate the model in such a way as to treat all goods (including productive factors as well as consumption goods) as symmetrical. Let us represent the goods by  $i = 1, \dots, n + r$ , and denote by an  $(n + r)$ -vector the outputs of all goods (a negative amount here represents an input).

The technology of the system is described by specifying the production possibility set  $Z$  as a subset of the  $(n + r)$ -vector space such that an output vector  $\mathbf{z}$  is technologically possible if and only if  $\mathbf{z} \in Z$ . The assumption corresponding to the laws of constant returns to scale and of the diminishing marginal rates of transformation will be that  $Z$  is a convex cone in the  $(n + r)$ -space.

The original setup of the production model can be represented by a production possibility set. Explicitly, if the technology is formulated in terms of  $\mathbf{f}(\mathbf{x})$  and  $\mathbf{g}(\mathbf{x})$ , the corresponding  $Z$  is defined by

$$Z = \left\{ \begin{pmatrix} \mathbf{f}(\mathbf{x}) \\ -\mathbf{g}(\mathbf{x}) \end{pmatrix} \mid \mathbf{x} \geq 0 \right\}.$$

On the other hand, if the technology is given in terms of  $Z$ , the corresponding  $\mathbf{f}(\mathbf{x})$  and  $\mathbf{g}(\mathbf{x})$  are defined by taking a parametric representation of  $Z$ , if this is possible, namely  $\mathbf{z} = \mathbf{z}(\mathbf{x})(\mathbf{x} \geq \mathbf{0})$ , and defining

$$\mathbf{f}(\mathbf{x}) = \langle z_1(\mathbf{x}), \dots, z_n(\mathbf{x}) \rangle, \quad \mathbf{g}(\mathbf{x}) = \langle -z_{n+1}(\mathbf{x}), \dots, -z_{n+r}(\mathbf{x}) \rangle.$$

Considered in terms of the production possibility sets, the production problem is a matter of determining and characterizing all efficient technologically possible production vectors. An instance of this will be examined in the following section.

A specific efficient production vector may be arrived at by taking account of market prices, as is done in (8.4.1).

*The classical Leontief production model.* The open Leontief production model corresponds in the above formulation to the case in which  $r = 1$  (labor) and

$$Z = \left\{ \begin{pmatrix} (\mathbf{I} - \mathbf{A})\mathbf{x} \\ (-1, \mathbf{x}) \end{pmatrix} \middle| \begin{array}{l} \mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle \\ \mathbf{x} \geq \mathbf{0} \end{array} \right\},$$

where  $\mathbf{A} = \|a_{ik}\|$  stands for the input-output coefficient matrix and  $\mathbf{l} = \langle l_1, l_2, \dots, l_n \rangle$  represents the labor coefficient vector. Here  $x_i$  may be interpreted directly as the amount of the  $i$ th product produced, and  $a_{ik}$  measures the amount of the  $i$ th product used in producing one unit of the  $k$ th product. The negative of the last component of  $Z$  evaluates the quantity of labor consumed for a given output vector.

**8.5 Efficient points of a Leontief-type model.** One of the main limitations of the Leontief model as described in the preceding section is that it associates only one activity with the production of each product. We shall see in this section that even if many processes are available for each commodity, nevertheless all “efficient” modes of production involve a single production process for each activity, provided that no joint production is allowed and only one primary factor is available. This is the famous Samuelson substitution theorem. We illustrate this concept in the following model.

Suppose that we have  $n$  industries engaged in the production of  $n$  commodities. Let  $a_{ij}$  be the amount of the  $j$ th product produced (or consumed) by the  $i$ th industry when it is operating at the unit level of activity.

We assume that  $a_{ij} \leq 0$  if  $i \neq j$ , but that  $a_{ii} > 0$ . A negative  $a_{ij}$  may be interpreted as material used up in the production process. We see that whereas each industry may use all commodities, it produces only one (this is what is meant by no joint production).

If we represent the total amount of available labor by unity, then the level of activity of the  $i$ th industry may be represented by  $x_i \geq 0$ , where

$x_i$  is the proportion of the total available labor employed in this industry. The net product of commodity  $j$  is  $\sum_i x_i a_{ij}$ . Suppose that the  $i$ th industry has available alternative processes  $a_{ij}^{\lambda_i}$  ( $\Lambda_i$  will index the set of processes available in the  $i$ th industry;  $\lambda_i \in \Lambda_i$ ); then the net product of commodity  $j$ , corresponding to a choice of production processes  $\lambda_1, \dots, \lambda_n$  for industries  $1, \dots, n$ , respectively, is  $\sum_i x_i a_{ij}^{\lambda_i}$ .

An alternative interpretation of this model is that the manufacturers vary the number of factories engaged in the production of a given commodity.

Let  $S_k$  ( $k = 1, \dots, n$ ) denote the set of process vectors available to the  $k$ th industry. For fixed  $k$ , the set  $S_k$  consists of all possible process vectors  $\mathbf{a}_k^{\lambda_k} = \langle a_{k1}^{\lambda_k}, \dots, a_{kn}^{\lambda_k} \rangle$ . We impose the requirement that  $S_k$  is convex and compact for each  $k$ ; this is a natural assumption in terms of the alternative interpretation and can be justified as reasonable for many situations. By the set  $S$  we mean the subset of  $E^n$  which is the convex hull of the sets  $S_1, \dots, S_n$  (see Section B.2).

The set  $\{\mathbf{A}^{\lambda}\} = \mathcal{A}$  of matrices representing possible processes consists of the matrices  $\|a_{ij}^{\lambda_i}\|$  of order  $n$ ; the  $i$ th row is an element of the set  $S_i$ .

A vector  $\mathbf{c}$  in  $E^n$  is called feasible if  $\mathbf{c} = \mathbf{x}\mathbf{A}$  for some  $\mathbf{A}$  in  $\mathcal{A}$  and  $\mathbf{x}$  such that  $x_i \geq 0$ ,  $\sum x_i = 1$ . (Vectors  $\mathbf{x}$  of this form will be called admissible.)

► **LEMMA 8.5.1.** The set of all possible feasible vectors coincides with the set  $S$ .

The verification of this assertion is direct and will be omitted.

A feasible point  $\mathbf{c} = \mathbf{x}^0 \mathbf{A}^{\lambda^0}$  is efficient if there exists no admissible  $\mathbf{x}$  and  $\mathbf{A}^{\lambda}$  in  $\mathcal{A}$  such that  $\mathbf{x}^0 \mathbf{A}^{\lambda^0} \leq \mathbf{x} \mathbf{A}^{\lambda}$  with strict inequality in at least one component.

The efficient point concept introduced here is identical to the concept of an efficient point in nonlinear programming theory (see Section 7.4).

Our present objective is to characterize all efficient points for this Leontief model. For this purpose the following lemma will be needed.

► **LEMMA 8.5.2.** Let  $\mathbf{p}_i$  denote vectors belonging to  $S_i$  ( $i = 1, \dots, n$ ). Form the plane generated by the  $n$  points  $\mathbf{p}_i$ , that is, the sets of all points  $\sum_i y_i \mathbf{p}_i$ ,  $\sum_i y_i = 1$ . If this plane intersects the interior of the first quadrant for  $\mathbf{y} \geq \mathbf{0}$ , then the intersection of the plane and the first quadrant is contained in the set  $\Gamma = \{\bar{\mathbf{p}} = \sum_i x_i \mathbf{p}_i | x_i \geq 0, \sum x_i = 1\}$ .

*Proof.* (1) Let  $\mathbf{A}^{\lambda}$  denote the process matrix constructed from the vectors  $\mathbf{p}_i$ . Suppose that  $x_i \geq 0$ ,  $\sum x_i = 1$ , and  $x_{i_0} = 0$  for some  $i_0$ . Then  $\mathbf{x}\mathbf{A}^{\lambda}$  will have at least one nonpositive component, namely  $i_0$ , because  $a_{i_0 j}^{\lambda_i} \leq 0$  for  $i \neq i_0$ . Thus if  $x_i \geq 0$ ,  $\sum x_i = 1$ , and  $x_{i_0} = 0$ , it follows that  $\mathbf{x}\mathbf{A}^{\lambda}$  does not belong to the interior of the first quadrant.

(2) The set of all  $\mathbf{y}$  such that  $\sum y_i = 1$  and  $\sum y_i p_i \gg 0$  is clearly convex. By hypothesis and by (1) there exists an  $\mathbf{x}^0$  such that  $x_i^0 > 0$ ,  $\sum x_i^0 = 1$ , and  $\sum x_i^0 p_i \gg 0$ . Suppose that there exists a  $\mathbf{y}$  such that  $\sum y_i = 1$  and  $\sum y_i p_i \gg 0$  but  $y_i < 0$  for some  $i$ . Then we can find a value  $\lambda$  ( $0 < \lambda < 1$ ) such that the nonnegative vector  $\lambda \mathbf{x}^0 + (1 - \lambda) \mathbf{y}$  has at least one component equal to zero. But it follows from (1) that  $\lambda \mathbf{x}^0 + (1 - \lambda) \mathbf{y} \gg 0$ , which is impossible. Hence  $y_i \geq 0$  for all  $i$ .

The validity of Lemma 8.5.2 depends in a vital way on the hypothesis. The plane must intersect the interior of the positive orthant; otherwise the conclusion of the lemma is false. The reader can readily supply counterexamples in  $E^3$ ; e.g., take

$$\mathbf{p}_1 = \langle \frac{1}{2}, -\frac{1}{2}, 0 \rangle, \quad \mathbf{p}_2 = \langle -\frac{1}{2}, \frac{1}{2}, 0 \rangle, \quad \mathbf{p}_3 = \langle 0, 0, 1 \rangle.$$

With the aid of Lemma 8.5.2 we are now able to prove the principal theorem of this section.

►THEOREM 8.5.1. If there exists a feasible point in the interior of the positive orthant, then the efficient points in the positive orthant belong to a hyperplane and every point of the intersection of the hyperplane with the positive orthant is efficient.

*Proof.* Let  $\mathbf{p}_0$  be the feasible point, with strictly positive coordinates, and let  $l$  be the ray from the origin through the point  $\mathbf{p}_0$ . For every set of  $n$  points  $\mathbf{p}_1, \dots, \mathbf{p}_n$  ( $\mathbf{p}_i \in S_i$ ), which determines a feasible point, construct the plane generated by these points. Consider the plane that intersects the ray  $l$  at the point of the positive orthant that is farthest from the origin. Let this plane be denoted by  $P$ . Then  $P$  is the desired hyperplane; its intersection  $P'$  with the positive orthant gives the set of efficient points in the positive orthant.

Suppose that this is not so; i.e., suppose that there exists a feasible point  $\mathbf{g}$  in the positive orthant which dominates ( $\geq$ , with  $>$  in at least one component) some point  $\mathbf{p}$  of  $P'$ . But,  $P$  intersects the coordinate axes in the points  $\langle a_1, 0, \dots, 0 \rangle, \langle 0, a_2, 0, \dots, 0 \rangle, \dots, \langle 0, \dots, 0, a_n \rangle$ ; and by Lemma 8.5.2 all these points are feasible points. The point  $\mathbf{g}$  also belongs to a plane  $G$  generated by a set of points  $\tilde{\mathbf{g}}_1, \dots, \tilde{\mathbf{g}}_n$  ( $\tilde{\mathbf{g}}_i \in S_i$ ), and  $G$  intersects the coordinate axes in the points  $\langle b_1, 0, \dots, 0 \rangle, \langle 0, b_2, 0, \dots, 0 \rangle, \dots, \langle 0, \dots, 0, b_n \rangle$ , which are feasible points by the lemma. Since  $\mathbf{g}$  dominates  $\mathbf{p}$ , at least one of the  $b_i$ 's must be greater than one of the  $a_i$ 's (say  $b_1$ ). The points  $\langle b_1, 0, \dots, 0 \rangle, \langle 0, a_2, 0, \dots, 0 \rangle, \dots, \langle 0, \dots, 0, a_n \rangle$  are all feasible; hence their convex hull is a subset of  $S$  in the positive orthant. The points of this subset dominate the points of  $P'$ , and in particular this subset intersects the ray at a point farther from the origin than  $P'$ . But this contradicts our construction of  $P'$ . Therefore  $P'$  is the set of efficient points in the positive orthant.

There are three interesting consequences of this theorem. If there is a combination of processes and labor which yields a strictly positive surplus for each substance, then:

- (1) There is one process for each industry which gives the nonnegative efficient points. Each industry need not substitute different processes to secure nonnegative efficient points. Each can always use the same process.
- (2) The set of nonnegative efficient points is partially independent of labor, in that there are many distributions of labor which yield nonnegative efficient points.
- (3) Prices are independent of demands, since the normal vector to the unique efficient hyperplane determines the same relative price system for all efficient modes of production.

†*Computational procedure for the Leontief model, and its connection with completely mixed games.* Theorem 8.5.1 shows that all efficient points of the positive orthant lie in a hyperplane, and the proof of the theorem also establishes that this hyperplane is constructed from a process matrix  $\mathbf{A} \in \mathcal{Q}$ . Since this plane is assumed to intersect the interior of the positive orthant, for this  $A$  there exists a vector  $\mathbf{x}^0$  such that  $\mathbf{x}^0 \mathbf{A} \gg \mathbf{0}$ .

Furthermore, the arguments of the theorem indicate that it is sufficient to find the matrix  $A$  for which  $xA$  is a maximum along any particular ray entering the first quadrant in  $E^n$ . We choose to maximize along the 45-degree line. Obviously, the problem is equivalent to determining the matrix  $A$  for which

$$\max_{A \in \mathcal{Q}} v(\mathbf{A})$$

is achieved, where  $v(\mathbf{A}) = \max_{\mathbf{x}} \min_j \sum_i x_i a_{ij}$ .

The quantity  $v(\mathbf{A})$  is recognized as the value of the game defined by the matrix  $\mathbf{A}$ . These matrix games are of the form referred to earlier as Minkowski-Leontief games (see Section 2.5). An important property they possess is stated in the following lemma.

►LEMMA 8.5.3. If  $\mathbf{A} \in \mathcal{Q}$  and there exists a vector  $\mathbf{x}^0$  such that  $\mathbf{x}^0 \mathbf{A} \gg \mathbf{0}$ , then  $\mathbf{A}$  is completely mixed.

*Proof.* The relation  $\mathbf{x}^0 \mathbf{A} \gg \mathbf{0}$  implies that  $v(\mathbf{A}) > 0$ . If any optimal  $x$  strategy had a zero component, the yield for the corresponding column would be negative, hence less than  $v$ . Thus every optimal strategy for Player I has all components strictly positive.

It follows from Theorem 2.4.1 that the kernel of any pair of optimal strategies must be the matrix  $\mathbf{A}$ , and the optimal  $y$  strategy must be unique.

---

† The remainder of this section may be passed over on first reading without loss of continuity.

The dimension relations also show that  $\mathbf{y}$  must have all components positive.

Theorem 2.4.1 shows also that

$$v(\mathbf{A}) = \frac{1}{(\mathbf{u}, \mathbf{A}^{-1}\mathbf{u})} = \frac{1}{\sum_{i,j} A_{ij}},$$

where  $\|A_{ij}\| = \mathbf{A}^{-1}$ .

Let  $\mathfrak{A} = \{\mathbf{A} \in \mathfrak{Q} | v(\mathbf{A}) > 0\}$ . To characterize the efficient points it is sufficient to find the matrix  $\mathbf{A} \in \mathfrak{A}$  for which  $v(\mathbf{A})$  is a maximum, or equivalently to find  $\mathbf{A}$  such that

$$\max_{\mathbf{A} \in \mathfrak{A}} \frac{1}{\sum_{i,j} A_{ij}}.$$

**\*8.6 The theory of consumer choice.** The general theory of equilibrium that describes the relationship of supply and demand to prices and other factors of the economy can be divided into two main parts: (1) production theory, which also embraces the theory of distribution of value (as embodied, for example, in the shadow price concept); and (2) consumption theory, which is concerned with individual buying preferences. In the previous sections we have reviewed the mathematical theory of production. In this section we concentrate on describing some of the problems arising in the mathematical theory of consumer choice. In Sections 8.7 and 8.8 we examine a global equilibrium model which recognizes simultaneously the production and consumption factors.

The theory of consumer choice is concerned essentially with the following question: Given a consumer with a limited budget and a definite set of preferences with respect to different commodity bundles, what quantities does he consume when confronted with given market prices for the various commodities? Another important question is this: By what rules does such a consumer's preference pattern change in response to variations in market prices or in his income?

Suppose that there are  $n$  commodities consumed by the consumer, and let the quantities of these commodities be represented by an  $n$ -vector  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , where  $x_i$  stands for the amount of the  $i$ th commodity consumed. We may assume, by taking an appropriate scale, that  $x_i \geq 0$  ( $i = 1, \dots, n$ ).

The consumer is assumed to have a preference relation  $P$  between vectors. That is to say, given any two vectors  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  and  $\mathbf{y} = \langle y_1, y_2, \dots, y_n \rangle$ , either  $\mathbf{x}$  is preferred to  $\mathbf{y}$  ( $\mathbf{x} P \mathbf{y}$ ), or  $\mathbf{y}$  is preferred to  $\mathbf{x}$  ( $\mathbf{y} P \mathbf{x}$ ), or neither is preferred to the other ( $\mathbf{x} P \mathbf{y}$  and  $\mathbf{y} P \mathbf{x}$ ), where

---

\*Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

$\overline{\mathbf{x} P \mathbf{y}}$  stands for the negation of  $\mathbf{x} P \mathbf{y}$ ). In this last case we shall also say that  $\mathbf{x}$  is indifferent to  $\mathbf{y}$  and write  $\mathbf{x} I \mathbf{y}$ . We do not claim transitivity for the indifference relation at present, although we shall see later that once postulates P-1 to P-5 below are accepted, the  $I$  relations indeed become transitive and reflexive. We postulate that  $P$  has the following properties:

- P-1. Irreflexivity:  $\overline{\mathbf{x} P \mathbf{x}}$  for any  $\mathbf{x} \geq 0$ .
- P-2. Transitivity:  $\mathbf{x} P \mathbf{y}$  and  $\mathbf{y} P \mathbf{z}$  imply  $\mathbf{x} P \mathbf{z}$ .
- P-3. Monotonicity;  $\mathbf{x} > \mathbf{y}$  implies  $\mathbf{x} P \mathbf{y}$ .
- P-4. Convexity:  $\mathbf{x} P \mathbf{z}$  or  $\mathbf{x} I \mathbf{z}$  and  $\mathbf{y} P \mathbf{z}$  or  $\mathbf{y} I \mathbf{z}$  with  $\mathbf{y} \neq \mathbf{x}$  imply  $[t\mathbf{x} + (1 - t)\mathbf{y}]P \mathbf{z}$  for any  $0 < t < 1$ .
- P-5. Continuity:  $P_x = \{\mathbf{y} | \mathbf{y} P \mathbf{x}\}$  and  $Q_x = \{\mathbf{y} | \mathbf{x} P \mathbf{y}\}$  are open sets for any  $\mathbf{x}$ .

If there is a numerical function  $U(\mathbf{x})$  ( $\mathbf{x} \geq 0$ ) such that

$$U(\mathbf{x}) > U(\mathbf{y}) \quad \text{if and only if } \mathbf{x} P \mathbf{y}, \quad (8.6.1)$$

(i.e.,  $U(\mathbf{x}) > U(\mathbf{y})$  and  $\mathbf{x} P \mathbf{y}$  are equivalent statements), then  $U$  is called a utility index of  $P$ . If  $U$  is an index and  $F(U)$  is a monotone-increasing function of  $U$ , then  $V = F(U)$  is also an index.

In order for the monotonicity and convexity assumptions to be satisfied by a utility index function  $U$ , it is necessary and sufficient that the set  $\{\mathbf{x} | U(\mathbf{x}) \geq c\}$  defined for each constant  $c$  be a strictly convex set [see Section B.4, property (i)] which includes with each  $\mathbf{x}$  all vectors  $\mathbf{y} > \mathbf{x}$ . The postulate of continuity is implied whenever  $U$  is continuous. Conversely,

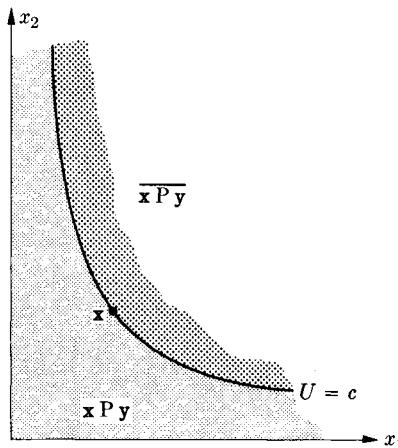


FIGURE 8-1

any real-valued function  $U$  with these properties defines a preference ordering by (8.6.1). Clearly  $\mathbf{x} \succ \mathbf{y}$  is synonymous with  $U(\mathbf{x}) = U(\mathbf{y})$ .

The preference relation induced by an index function is pictured in Fig. 8.1. The dotted region consists of all  $\mathbf{y}$  to which  $\mathbf{x}$  is preferred. This is an open set. The crosshatched region, which includes the boundary surface  $\{\mathbf{y} | U(\mathbf{y}) \geq c\}$ , represents all  $\mathbf{y}$  to which  $\mathbf{x}$  is not preferred. The same geometrical picture applies to the general preference relation  $P$  that satisfies postulates P-1 to P-5.

*Demand function for the preference relation  $P$ .* We shall assume that the consumer behaves in such a manner as to maximize his preference among all commodity vectors  $\mathbf{x}$  satisfying his budget constraint. More precisely, given market prices  $\mathbf{p} = \langle p_1, p_2, \dots, p_n \rangle$  and his income  $M$ , he chooses a vector  $\mathbf{x}$  for which

$$\mathbf{x} \geq \mathbf{0}, \quad (\mathbf{p}, \mathbf{x}) \leq M, \quad (8.6.2)$$

where  $(\mathbf{p}, \mathbf{x})$  evaluates the total cost to the consumer of the consumption vector when the market prices are given by  $\mathbf{p}$ , and

$$\mathbf{x} \succ \mathbf{y} \quad \text{for any } \mathbf{y} \geq \mathbf{0}, \quad (\mathbf{p}, \mathbf{y}) \leq M, \quad (8.6.3)$$

where  $\mathbf{p} \gg \mathbf{0}$  and  $M \geq 0$ .

By postulates P-1 through P-5, it is easily seen that the vector  $\mathbf{x}$  satisfying (8.6.2) and (8.6.3) always exists and is uniquely determined by  $\mathbf{p}$  and  $M$ .

The proof of uniqueness is an immediate consequence of transitivity and irreflexivity. Since the proof of existence is slightly more complicated, we present a formal proof.

For each  $\mathbf{x}$  satisfying  $(\mathbf{p}, \mathbf{x}) \leq M$ , we consider the set

$$B_x = \{\mathbf{y} | (\mathbf{p}, \mathbf{y}) \leq M, \mathbf{x} \succ \mathbf{y}\}. \quad (8.6.4)$$

We prove first that  $\Gamma = B_{x_1} \cap B_{x_2} \cap \dots \cap B_{x_n}$  is nonvoid for any finite  $\mathbf{x}_i$ . To this end, let the set  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  be separated into two groups  $X^1$  and  $X^2$ , where some member of  $X^1$  is preferred to each member of  $X^2$ , and where  $\mathbf{x}_i \succ \mathbf{x}_j$  for each  $\mathbf{x}_i$  and  $\mathbf{x}_j$  in  $X^1$ . Such a decomposition is possible with  $X^1$  nonvoid by virtue of the irreflexivity and transitivity postulated. All one has to do is to start systematically with  $\mathbf{x}_1$  and  $\mathbf{x}_2$  and compare them. If one of them is preferred to the other, the dominated member is assigned to  $X^2$ . We then compare the remaining member with  $\mathbf{x}_3$ , etc.

Now suppose that the  $\mathbf{x}_i$  are labeled such that  $\mathbf{x}_1, \dots, \mathbf{x}_k$  belong to  $X^1$  and  $\mathbf{x}_{k+1}, \dots, \mathbf{x}_n$  belong to  $X^2$ . We prove that  $(\mathbf{x}_1 + \dots + \mathbf{x}_k)/k$  is in  $\Gamma$ . By P-4,

$$\frac{\mathbf{x}_1 + \dots + \mathbf{x}_k}{k} \succ \mathbf{x}_i \quad (i = 1, \dots, k),$$

and by transitivity and the definition of  $X^2$ ,

$$\frac{\mathbf{x}_1 + \cdots + \mathbf{x}_k}{k} P \mathbf{x}_j \quad (j = 1, \dots, n).$$

Furthermore,

$$\mathbf{x}_j P \frac{\mathbf{x}_1 + \cdots + \mathbf{x}_k}{k},$$

for otherwise  $P-1$  would be contradicted for the vector  $(\mathbf{x}_1 + \cdots + \mathbf{x}_k)/k$ . Since each of the sets  $B_x$  is a closed subset (P-5) of a compact set  $T = \{y|(\mathbf{p}, y) \leq M, \mathbf{p} \gg \mathbf{0}\}$  and possesses the finite intersection property, it follows that

$$\Lambda = \bigcap_{x \in T} B_x$$

is nonvoid. Let  $\mathbf{x}^0$  be in  $\Lambda$ ; then  $\mathbf{x}^0$  satisfies the budget constraint. Moreover,  $\mathbf{x}^0 P y$  for all  $y \neq \mathbf{x}^0$ . If we suppose the contrary, then  $\mathbf{x}^0 P \overline{y^0}$  for some  $y^0$  in  $T$ , and  $\overline{y^0 P \mathbf{x}^0}$  by the construction of  $\Lambda$ . Hence  $y^0 I \mathbf{x}^0$ . Also  $\mathbf{x}^0 I \mathbf{x}^0$ , and hence  $[t\mathbf{x}^0 + (1-t)y^0] P \mathbf{x}^0$  for  $0 < t < 1$  by P-4, contradicting the choice of  $\mathbf{x}^0$ .

We may write

$$\mathbf{x}^0 = \mathbf{f}(\mathbf{p}, M),$$

and  $f$  will be called the demand function derived from the preference relation  $P$ .

The demand function may also be characterized as the vector  $\mathbf{x}^0$  satisfying

$$\mathbf{x}^0 \geq \mathbf{0}, \quad (\mathbf{p}, \mathbf{x}^0) = M, \quad \mathbf{x}^0 P \mathbf{x} \quad (8.6.5)$$

for all  $\mathbf{x}$  such that  $(\mathbf{p}, \mathbf{x}) < M$  when  $\mathbf{p} \gg \mathbf{0}$  and  $M > 0$ .

*Proof.* If  $\mathbf{x}^0$  is the demand function for the parameters  $\mathbf{p}$  and  $M$  and  $(\mathbf{p}, \mathbf{x}^0) < M$ , then we may increase all the components of  $\mathbf{x}^0$  while maintaining  $(\mathbf{p}, \mathbf{x}) \leq M$ . But by P-3 the property (8.6.3) is contradicted.

Conversely, suppose that  $\mathbf{x}^0$  satisfies (8.6.5). In this case we show that  $\mathbf{x}^0$  is the demand function of the preference relation for the parameters  $\mathbf{p}$  and  $M$ . Suppose that the demand for  $\mathbf{p}$  and  $M$  is  $\mathbf{y}^0 \neq \mathbf{x}^0$ . Let

$$Q_{x^0} = \{y|y P \mathbf{x}^0\}.$$

This set is open (by P-5) and includes  $\mathbf{y}^0$ . Hence there exists a  $\mathbf{y}^0$  with the property  $\mathbf{y} P \mathbf{x}^0$  and  $(\mathbf{p}, \mathbf{y}) < M$ , which is impossible by (8.6.5). Hence  $\mathbf{y}^0$  agrees with  $\mathbf{x}^0$ , as was to be proved.

It will be helpful for future discussions to diagram the position of  $\mathbf{x} = \mathbf{f}(\mathbf{p}, M)$ ; this is done in Fig. 8.2.

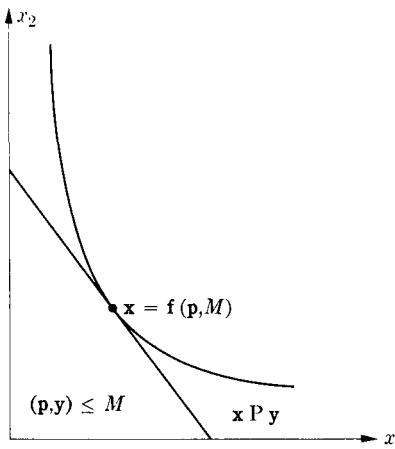


FIGURE 8-2

If the preference relation is described by a utility index  $U$ , then  $\mathbf{x}^0$  is the vector for which  $U(\mathbf{x})$  is maximized with respect to all  $\mathbf{x}$  satisfying the restrictions  $(\mathbf{p}, \mathbf{x}) \leq M$ .

In view of (8.6.5), it is easily shown that  $\mathbf{f}$  is continuous at  $\mathbf{p} \gg 0$  and  $M > 0$ . In general, by a demand function  $\mathbf{x} = \mathbf{f}(\mathbf{p}, M)$  will be meant a function  $\mathbf{f}$ , defined on  $\mathbf{p} \gg 0$  and  $M \geq 0$ , such that

$$\text{D-1. } \mathbf{f}(\mathbf{p}, M) \geq 0, (\mathbf{p}, \mathbf{f}(\mathbf{p}, M)) = M.$$

$$\text{D-2. } \mathbf{f} \text{ is continuous at } \mathbf{p} \gg 0, M > 0.$$

In what follows we shall make two further assumptions:

D-3. For any  $\mathbf{x} \geq 0$ , there exist  $\mathbf{p} \gg 0$  and  $M \geq 0$  such that  $\mathbf{x} = \mathbf{f}(\mathbf{p}, M)$ .

D-4.  $\mathbf{f}(\mathbf{p}, M)$  satisfies a Lipschitz condition with respect to  $M > 0$ ; i.e., for any  $M > 0$ , there exist  $K > 0$  and  $\delta > 0$  such that

$$|\mathbf{f}(\mathbf{p}, M_1) - \mathbf{f}(\mathbf{p}, M)| \leq K|M_1 - M|$$

whenever  $|M_1 - M| < \delta$ .

If  $P$  is derived from a utility index  $U$ , where  $\{\mathbf{x}|U(\mathbf{x}) \geq c\}$  is a strictly convex set with property P-3 and  $U$  is sufficiently differentiable, then it is easily verified that D-3 and D-4 are satisfied.

*The axiom of revealed preference.* With the aid of a given demand function  $\mathbf{x} = \mathbf{f}(\mathbf{p}, M)$  fulfilling conditions D-1 through D-4, we now define the concept of "revealed preference." It is said that vector  $\mathbf{x}^0 = \mathbf{f}(\mathbf{p}^0, M^0)$  is revealed-preferred to vector  $\mathbf{x}^1 = \mathbf{f}(\mathbf{p}^1, M^1)$  if

$$(\mathbf{p}^0, \mathbf{x}^0) \geq (\mathbf{p}^0, \mathbf{x}^1) \quad (\mathbf{x}^0 \neq \mathbf{x}^1).$$

This relationship will be designated by

$$\mathbf{x}^0 R \mathbf{x}^1;$$

that is,  $\mathbf{x}^0 R \mathbf{x}^1$  if  $\mathbf{x}^1$  belongs to the budget constraint set over which  $\mathbf{x}^0$  is demanded. Naturally, to be consistent with the meaning of demand functions, it is impossible for  $\mathbf{x}^0$  to be an admissible consumption vector for the budget constraint satisfied by  $\mathbf{x}^1$ .

We also define relations  $R^s$  and  $R^*$  as follows:  $\mathbf{x} R^s \mathbf{y}$  if there exist vectors  $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{s-1}$  such that  $\mathbf{x} R \mathbf{x}^1, \mathbf{x}^1 R \mathbf{x}^2, \dots, \mathbf{x}^{s-1} R \mathbf{y}$ ; and  $\mathbf{x} R^* \mathbf{y}$  if  $\mathbf{x} R^s \mathbf{y}$  for some integer  $s$ . It is readily shown that the relation  $R^*$  is transitive, monotonic, and continuous. Moreover, whenever  $\mathbf{x} = \mathbf{f}(\mathbf{p}, M)$  is derived from a preference relation  $P$ , the relation  $R^*$  is irreflexive: for since  $\mathbf{x} R \mathbf{y}$  implies  $\mathbf{x} P \mathbf{y}$ , it follows that  $\mathbf{x} R^* \mathbf{y}$  implies  $\mathbf{x} P \mathbf{y}$  by the transitivity of  $P$ ; hence the irreflexivity of  $R^*$  is a consequence of that of  $P$ .

Furthermore, if the induced demand function satisfies D-3 and D-4, then relation  $R^*$  coincides with  $P$ :

$$R^* = P.$$

The proof of this is involved and will be omitted (see the notes at the close of this chapter). However, for the case of  $n = 2$  the orientation of the unit circle can be exploited and the proof is not difficult (see Problem 19).

It can also be shown that if the demand function  $\mathbf{f}(\mathbf{p}, M)$  satisfies conditions D-1 through D-4 and  $R^*$  is irreflexive, then  $R^*$  is a preference relation satisfying postulates P-1 through P-5, and  $\mathbf{f}(\mathbf{p}, M)$  is the demand function derived from  $R^*$ .

The irreflexivity of  $R^*$  may be characterized as follows: For any integer  $s$ , the relations

$$(\mathbf{p}^0, \mathbf{x}^0) \geq (\mathbf{p}^0, \mathbf{x}^1), (\mathbf{p}^1, \mathbf{x}^1) \geq (\mathbf{p}^1, \mathbf{x}^2), \dots, (\mathbf{p}^{s-1}, \mathbf{x}^{s-1}) \geq (\mathbf{p}^{s-1}, \mathbf{x}^s) \\ (\mathbf{x}^i \neq \mathbf{x}^j \text{ for some } i, j)$$

imply

$$(\mathbf{p}^s, \mathbf{x}^0) > (\mathbf{p}^s, \mathbf{x}^s), \quad (8.6.6)$$

where

$$\mathbf{x}^t = \mathbf{f}(\mathbf{p}^t, M^t) \quad (t = 0, 1, \dots, s).$$

This characterization is usually called the *strong axiom of revealed preference*. A related concept, the *weak axiom of revealed preference*, is formulated as follows:

$$(\mathbf{p}^0, \mathbf{x}^0) \geq (\mathbf{p}^0, \mathbf{x}^1) \quad (\mathbf{x}^0 \neq \mathbf{x}^1) \quad \text{implies} \quad (\mathbf{p}^1, \mathbf{x}^0) > (\mathbf{p}^1, \mathbf{x}^1). \quad (8.6.7)$$

The strong axiom obviously implies the weak axiom. We shall have occasion to apply the concept of revealed preference in our later discussion of stability (see Section 9.4).

*Compensated income function for price  $\mathbf{p}$ .* As an aid to our investigation of the properties of the demand function  $\mathbf{f}(\mathbf{p}, M)$  derived from the preference relation  $P$ , we introduce the notion of "compensated income change."

Let  $\mathbf{p}^0 \gg \mathbf{0}$ ,  $M^0 > 0$ , and  $\mathbf{x}^0 = \mathbf{f}(\mathbf{p}^0, M^0)$  be given. We define the compensated income  $M(\mathbf{p})$  for price  $\mathbf{p}$  by

$$M(\mathbf{p}) = \inf \{M \mid \mathbf{f}(\mathbf{p}, M) \geq \mathbf{x}^0\},$$

or, equivalently,

$$M(\mathbf{p}) = \sup \{M \mid \mathbf{x}^0 \geq \mathbf{f}(\mathbf{p}, M)\}.$$

Figure 8.3 helps to indicate the meaning of  $M(\mathbf{p})$ . For a given utility value that corresponds to the utility of  $\mathbf{x}^0$ , the compensated income  $M(\mathbf{p})$  can be interpreted as the minimum budget for a given price vector which will purchase the same utility as  $\mathbf{x}^0$ .

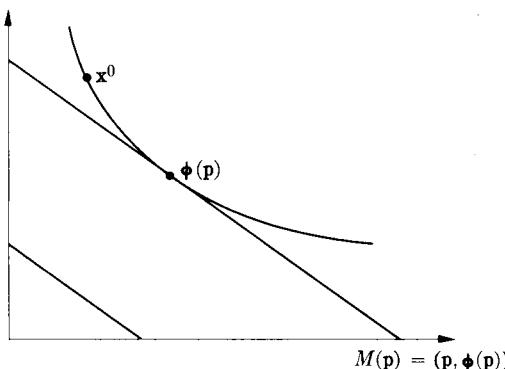


FIGURE 8-3

It is easy to show (by appealing to the geometrical description of  $M(\mathbf{p})$  as indicated in Fig. 8.3) that  $M(\mathbf{p})$  is a concave function of  $\mathbf{p}$ , and therefore that it is continuous at  $\mathbf{p} \gg \mathbf{0}$  and differentiable almost everywhere.

We assume for simplicity of exposition that  $M(\mathbf{p})$  is differentiable everywhere. We may also regard  $M(\mathbf{p})$  as a support function of a convex set; explicitly,

$$M(\mathbf{p}) = \min \{ \mathbf{p} \cdot \mathbf{x} \},$$

where the minimum is extended over all  $\mathbf{x}$  for which  $\overline{\mathbf{x}_0 \mathbf{P} \mathbf{x}}$ .  $M(\mathbf{p})$  is a concave function, being a minimum of linear functions. Let

$$\phi(\mathbf{p}) = \mathbf{f}[\mathbf{p}, M(\mathbf{p})]; \quad (8.6.8)$$

then

$$(\mathbf{p}, \phi(\mathbf{p})) = M(\mathbf{p}) \quad (8.6.9)$$

and

$$M(\mathbf{p}^0) = M^0, \quad \mathbf{x}^0 = \phi(\mathbf{p}^0).$$

According to the budget constraint  $M(\mathbf{p})$ , the vector  $\phi(\mathbf{p})$  is clearly the preferred consumption vector with respect to all  $\mathbf{x}$  meeting the condition  $(\mathbf{p}, \mathbf{x}) \leq M(\mathbf{p})$ .

It is clear from the geometry of Fig. 8.3 and from the definition of  $\phi(\mathbf{p})$  and  $M(\mathbf{p})$  (a rigorous proof is easy) that the function  $\sum \tilde{p}_i \phi_i(\mathbf{p}) = g(\mathbf{p})$ , with  $\tilde{\mathbf{p}} \gg \mathbf{0}$  as a parameter, is minimized as a function of  $\mathbf{p}$  uniquely at  $\mathbf{p} = \tilde{\mathbf{p}}$ . Hence the usual necessary condition required of a minimum value of a function yields the relation

$$\sum_{i=1}^n p_i \frac{\partial \phi_i}{\partial p_j} = 0 \quad (j = 1, 2, \dots, n). \quad (8.6.10)$$

From (8.6.10), we have

$$\left( \frac{\partial M(\mathbf{p})}{\partial p_j} \right)_j = \phi_j(\mathbf{p}) + \sum_{i=1}^n p_i \frac{\partial \phi_i}{\partial p_j} = \phi_j(\mathbf{p}). \quad (8.6.11)$$

All these manipulations are performed while keeping the utility level constant and equal to that of  $\mathbf{x}^0$ .

We shall define

$$K_{ij}(\mathbf{p}^0) = \left( \frac{\partial \phi_i}{\partial p_j} \right)_{p^0} = \left( \frac{\partial f_i}{\partial p_j} \right)_{U(x)=U(x^0)} \quad (8.6.12)$$

as the residual variability of the  $i$ th commodity with respect to the compensated price change of the  $j$ th commodity. In other words,  $K_{ij}$  is a measure of the change of desire for consumption of the  $i$ th commodity as the price of the  $j$ th commodity is varied, always keeping the same utility level. The idea behind the study of the function  $K_{ij}(\mathbf{p}^0)$  is as follows. Suppose that we attempt to assign meaning to the discrepancy of two distinct demand vectors,  $\mathbf{x}^0 = \mathbf{f}(\mathbf{p}^0, M^0)$  and  $\mathbf{x}' = \mathbf{f}(\mathbf{p}^1, M^1)$ . The change of  $\mathbf{x}^0$  to  $\mathbf{x}'$  can be decomposed into two components: (1) the change due to the change of prices in  $\mathbf{x}^0$ , keeping the same utility level; (2) the change for a fixed price resulting from a new budget size.  $K_{ij}(\mathbf{p}^0)$ , often referred to as the Slutsky term, can be regarded as a measure of the first change.

By (8.6.8) and (8.6.11), we have

$$\frac{\partial \phi_j}{\partial p_k} = \frac{\partial f_j}{\partial p_k} + \frac{\partial f_j}{\partial M} \frac{\partial M}{\partial p_k} = \frac{\partial f_j}{\partial p_k} + \frac{\partial f_j}{\partial M} \phi_k.$$

Evaluating (8.6.12) at  $\mathbf{p} = \mathbf{p}^0$ , we obtain

$$K_{jk}(\mathbf{p}^0) = \frac{\partial x_j^0}{\partial p_k} + \frac{\partial x_j^0}{\partial M} x_k^0,$$

where  $\mathbf{x}^0 = \mathbf{f}(\mathbf{p}^0, M^0)$ . Thus, if  $p_0$  is the only price that changes, the change in demand can be written as the difference between the Slutsky term and the term measuring the effective loss in income due to the change in price.

**THEOREM 8.6.1.**  $K_{jk}$  satisfies

$$K_{jk} = K_{kj} \quad (j, k = 1, \dots, n), \quad (8.6.13)$$

$$\sum_{i,j=1}^n K_{ij} h_i h_j \leq 0 \quad \text{for any real vector } \langle h_1, \dots, h_n \rangle, \quad (8.6.14)$$

$$\sum_j p_j^0 K_{jk}(\mathbf{p}^0) = 0 \quad (k = 1, \dots, n). \quad (8.6.15)$$

*Proof.* Equation (8.6.13) is easily implied by (8.6.11). Explicitly,

$$K_{jk}(\mathbf{p}^0) = \left( \frac{\partial \phi_j}{\partial p_k} \right)_0 = \left( \frac{\partial^2 M(p)}{\partial p_j \partial p_k} \right)_0 = \left( \frac{\partial^2 M(p)}{\partial p_k \partial p_j} \right)_0 = \left( \frac{\partial \phi_k}{\partial p_j} \right)_0 = K_{kj}(\mathbf{p}^0). \quad (8.6.16)$$

Equation (8.6.15) has already been proved in (8.6.10). Relation (8.6.14) is implied by the fact that  $M(\mathbf{p})$  is concave and by (8.6.11) and (8.6.12) (see Section B.4).

**8.7 Nonlinear models of equilibrium.** The state of an economic system under conditions of competition at any point in time can be formulated as the solution of a system of inequality relations expressing the demand for goods by consumers, the supply of goods by producers, and the equilibrium condition that supply exceeds demand in every market, it being assumed that if the supply of any commodity is overabundant, the price of that commodity is zero. From the study of the existence of equilibrium solutions, we acquire insight into descriptive and normative economic theory.

If competitive economic models are a reasonably accurate description of reality, then the fact that these models have solutions is one verification of their possible usefulness. Of equal importance is the relation of the existence of competitive-equilibrium solutions to the problem of welfare economics (see Section 9.1).

In this and the following section we discuss a series of similar nonoverlapping models describing phenomena of equilibrium and exchange. The key mathematical tool used in establishing the existence of solutions for

each of these models is some form of the Kakutani fixed-point theorem, with additional slight trimmings.

The linear equations of the Leontief models of Sections 8.1 and 8.3 can be interpreted as expressing an equilibrium of exchange among various activities. In order to bring the model into closer conformity with reality, we must abandon the assumption of "linearity," i.e., the condition that the numbers  $a_{ij}$  are constants independent of  $X_i$ . The study of certain basic nonlinear models will be the subject of the remainder of this section.

By way of introduction, we first consider the following model. Suppose that we have  $n$  producers, and that each producer  $P_i$  produces some commodity  $G_i$ . Let  $f_{ij}(x)$  ( $i \neq j$ ) represent the amount of money that  $P_i$  spends on  $G_j$  when his income is  $x$ . The functions  $f_{ij}(x)$  may be regarded as demand functions in the sense of Section 8.6. If each producer spends his entire income in buying goods from the other producers, then

$$x = \sum_{j=1}^n f_{ij}(x) \quad (8.7.1)$$

for each choice of  $i$  and positive  $x$  value. The function  $f_{ii}(x)$  is defined for convenience to be identically zero. The economic law states that the income to the  $i$ th producer  $x_i$  is determined such that the total amount of each commodity sold by a producer must equal the total amount of that commodity bought by the other producers. Mathematically, we seek to find values  $x_i$  which satisfy the system of equations

$$x_i = \sum_{i=1}^n f_{ij}(x_i). \quad (8.7.2)$$

This problem is a nonlinear analog of the problems of the closed Leontief model (see Theorem 8.3.1).

► **THEOREM 8.7.1.** If the nonnegative functions  $f_{ij}$  are continuous, and if (8.7.1) is satisfied for all nonnegative numbers, then there exist nonnegative numbers  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$  which satisfy (8.7.2).

*Proof.* Let  $S$  be the set of all nonnegative vectors  $\mathbf{y}$ , with

$$\sum_{i=1}^n y_i = 1.$$

We construct a continuous mapping  $T$  of  $S$  into itself. Define  $T$  such that the  $j$ th component of  $T\mathbf{y}$  is

$$(Ty)_j = \sum_{i=1}^n f_{ij}(\mathbf{y}).$$

It follows from (8.7.1) that  $T\mathbf{y}$  is in  $S$  whenever  $\mathbf{y}$  is chosen from  $S$ . The

mapping  $T$  is continuous, and hence by the Brouwer fixed-point theorem (see Section C.2 of the Appendix) there exists a vector  $\bar{\mathbf{x}}$  in  $S$  such that  $T\bar{\mathbf{x}} = \bar{\mathbf{x}}$ . The vector  $\bar{\mathbf{x}}$  obviously is a solution to the problem.

*A model with variable prices.* In the models considered up to now all prices were taken as fixed, and in order that supply should equal demand each producer was forced to control the volume of his output. We shall now take the opposite point of view and attempt to balance supply and demand by shifting prices rather than volume. To this end, we ask the question: Does there exist a set of prices such that the value of the goods sold by each producer is equal to the value of the goods he buys?

According to the classical law of supply and demand, the price of a commodity rises when the demand for that commodity exceeds the available supply and drops when the supply exceeds the demand. Economic equilibrium can therefore be achieved only when prices are such that supply and demand for each commodity are equal, taking into account the variability of prices. Let  $a_i$  be the amount of  $G_i$  produced by  $P_i$  in some fixed period of time, and let  $p_i$  be the price of one unit of  $G_i$ . If each producer sells his entire output, then his income will be  $p_i a_i$ . Let  $D_{ij}(p_1, p_2, \dots, p_n) = x_{ij}$  represent the amount of  $G_j$  demanded by  $P_i$ . The functions  $D_{ij}$  logically depend on  $a_i$  in the sense that the  $p_i$  are dependent on  $a_i$ . The assumption that each producer spends his entire income means that

$$p_i a_i = \sum_j p_j D_{ij}(\mathbf{p}) \quad (i = 1, 2, \dots, n). \quad (8.7.3)$$

(Set  $D_{ii} = 0$  by definition.)

We assume that (8.7.3) holds for any price vector  $\mathbf{p}$ . The total demand of goods  $G_j$  is  $\sum_i D_{ij}(\mathbf{p})$ , and the total supply is  $a_j$ . The condition for equilibrium is

$$a_j - \sum_i D_{ij}(\mathbf{p}) = 0. \quad (8.7.4)$$

The problem is: Does there exist for prescribed  $a_i$  and demand functions  $D$  which satisfy (8.7.3) a nonnegative solution  $\mathbf{p}$  of (8.7.4)? Unfortunately such a solution does not always exist. It is conceivable, for instance, that one of the producers may be supplying a commodity which none of the others want, in which case the left-hand side of (8.7.4) becomes positive. This consideration leads us to the following theorem.

► **THEOREM 8.7.2.** If the nonnegative function  $D_{ij}(\mathbf{p})$  satisfies (8.7.3) for every positive vector  $\mathbf{p}$  and  $a_i > 0$  for all  $i$ , then there exists  $\langle p_1^*, p_2^*, \dots, p_n^* \rangle$  such that

$$a_j - \sum_i D_{ij}(\mathbf{p}^*) \geq 0; \quad (8.7.5)$$

and if equality does not hold in the  $j_0$ th equation, then  $p_{j_0}^* = 0$ .

(The theorem states that there always exist prices for which supply is at least equal to demand, and that only goods which are oversupplied have price zero. This modification of the law of supply and demand is well known to economists. In this connection, see also Theorem 5.4.3.)

*Proof.* A real value  $\mu$  is called admissible if there exists a positive vector  $\mathbf{p} = \langle p_1, \dots, p_n \rangle$  such that  $\sum p_i = 1$  and

$$\mu a_j \geq F_j(\mathbf{p}), \quad (8.7.6)$$

where

$$F_j(\mathbf{p}) = \sum_i D_{ij}(\mathbf{p}) \quad (j = 1, \dots, n).$$

The set of all such normalized vectors  $\mathbf{p}$  will be denoted by  $S$ . The set  $V$  of all admissible  $\mu$  is clearly nonvoid and bounded below. Let  $\mu_0 = \inf \nu \mu$ ; then from the compactness and continuity of  $D_{ij}$  it follows that  $\mu_0$  is also admissible. Since  $\mu_0 - \epsilon$  is not admissible, for every  $\mathbf{p}$  in  $S$  ( $S$  = the set of all nonnegative vectors whose components sum to one) there exists at least one vector  $\mathbf{y}$  in  $S$  such that

$$((\mu_0 - \epsilon)\mathbf{a} - \mathbf{F}(\mathbf{p}), \mathbf{y}) \leq 0. \quad (8.7.7)$$

The set  $S(\mathbf{p})$  of all such  $\mathbf{y}$  in  $S$  satisfying (8.7.7) is convex, closed, and nonvoid. As a function of  $\mathbf{p}$ , the sets  $S(\mathbf{p})$  are continuous in the usual sense of point-to-set mapping (see Section C.2).

Since the requirements of the Kakutani fixed-point theorem are fulfilled, it follows that there exists an element  $\mathbf{p}^0$  in  $S$  such that

$$((\mu_0 - \epsilon)\mathbf{a} - \mathbf{F}(\mathbf{p}^0), \mathbf{p}^0) \leq 0. \quad (8.7.8)$$

This last equation, in conjunction with (8.7.3), implies that  $\mu_0 - \epsilon \leq 1$ , and hence  $\mu_0 \leq 1$ . If we multiply the  $j$ th equation of (8.7.6) by  $p_j$  and sum, it follows from (8.7.3) that  $\mu \geq 1$  for any admissible  $\mu$ . Therefore  $\mu_0 \geq 1$ , and thus  $\mu_0 = 1$ . Let  $\mathbf{p}^*$  be the vector corresponding to  $\mu_0 = 1$ . Then

$$a_j \geq \sum_{i=1}^n D_{ij}(\mathbf{p}^*) \quad (j = 1, \dots, n).$$

Since  $(\mathbf{a}_j, \mathbf{p}^*) - (\mathbf{F}(\mathbf{p}^*), \mathbf{p}^*) = 0$  by (8.7.3), it follows that  $p_j^* = 0$  whenever

$$a_j > \sum_{i=1}^n D_{ij}(\mathbf{p}^*).$$

*A model with variable prices and volumes.* In our next model both the amount and the type of goods supplied vary with prices. For instance, the quality of services supplied will often depend on the remuneration

received by the supplier. Consequently, we shall assume the existence of supply functions  $S_{ij}(\mathbf{p})$  ( $i \neq j$ ) which represent the amounts of  $G_j$  which  $P_i$  will supply at prices  $p_1, \dots, p_n$ . The demand functions  $D_{ij}(\mathbf{p})$  are specified as previously. The equation

$$\sum_{j=1}^n p_j S_{ij}(\mathbf{p}) = \sum_{j=1}^n p_j D_{ij}(\mathbf{p}) \quad (8.7.9)$$

(again  $S_{ii}(\mathbf{p})$  is defined to be identically zero) is the analog of (8.7.3). The law of supply and demand leads to the relation

$$\sum_i S_{ij}(\mathbf{p}) \geq \sum_i D_{ij}(\mathbf{p}). \quad (8.7.10)$$

We seek to find a price vector  $\mathbf{p}$  which satisfies (8.7.10), and such that the price of any commodity for which the supply exceeds the demand is zero.

► **THEOREM 8.7.3.** Let  $S_{ij}(\mathbf{p})$  and  $D_{ij}(\mathbf{p})$  be continuous nonnegative functions such that (8.7.9) is satisfied for each  $\mathbf{p}$  and  $\sum_i S_{ij}(\mathbf{p})$  is positive for each  $j$ . Then there exists a nonnegative vector

$$\mathbf{p}^* = \langle p_1^*, \dots, p_n^* \rangle \quad (\sum p_i^* = 1)$$

which satisfies (8.7.10), and in addition

$$\sum_j p_j^* \sum_i [S_{ij}(\mathbf{p}^*) - D_{ij}(\mathbf{p}^*)] = 0.$$

The proof parallels that of Theorem 8.7.2 and is therefore omitted.

† *A general theorem of Wald on the equilibrium equations.* Let

$$R_1, R_2, \dots, R_m$$

be the factors of production which in different combinations can be used to produce  $n$  different products  $S_1, S_2, \dots, S_n$ . Specifically,  $a_{ij}$  units of  $R_i$  are needed to produce one unit of  $S_j$ . Suppose that the producer has available  $r_1$  units of  $R_1$ ,  $r_2$  units of  $R_2, \dots, r_m$  units of  $R_m$ . If the price of a unit of  $S_j$  is  $f_j(s_1, \dots, s_n)$  when  $s_i$  units of  $S_i$  ( $i = 1, \dots, n$ ) are produced, then the equations for the unknown prices  $p_i$  for a unit of  $R_i$  and the unknown quantities  $s_j$  are

$$r_i = \sum_{j=1}^n a_{ij} s_j \quad (i = 1, \dots, m) \quad (8.7.11)$$

---

† The remainder of this section is of more special nature and may be omitted on first reading without loss of continuity.

and

$$\sigma_j = \sum_{i=1}^m a_{ij} p_i \quad (j = 1, \dots, n), \quad (8.7.12)$$

where  $\sigma_j = f_j(s_1, s_2, \dots, s_n)$ .

The equations as they stand need not have economically meaningful solutions. It has accordingly been suggested that (8.7.11) be replaced by

$$r_i \geq \sum_{j=1}^n a_{ij} s_j, \quad (8.7.13)$$

with the convention that wherever strict inequality holds, the corresponding price  $p_i$  is zero. This conforms with the general economic dictum that factors of production in excess of need should be priced zero (see Theorems 8.7.2 and 8.7.3).

Equation (8.7.12), which states that the production cost of a commodity equals its price, is retained without change. The general cost functions  $f_j$  are assumed to be continuous and nonnegative, and to satisfy the following condition:

(A) If a sequence of nonnegative  $s_1^k, \dots, s_n^k$  tends to  $s_1, \dots, s_n$ , respectively with  $s_j^k \neq 0$  for every  $k$  and  $s_j = 0$ , then

$$\lim_{k \rightarrow \infty} f_j(s_1^k, \dots, s_n^k) = \infty \quad (j = 1, \dots, n; k = 1, 2, \dots),$$

and otherwise  $f_j$  remains bounded.

This assumption is hard to justify on economic grounds. It is necessary, however, for the mathematical analysis which follows. One possible argument in its behalf is in terms of the existence of a fixed setup cost associated with producing.

- (B)  $r_i > 0$  ( $i = 1, \dots, m$ ).
- (C)  $a_{ij} \geq 0$  for all  $i, j$ , and for each  $j$  there exists at least one  $i$  such that  $a_{ij} > 0$ .

Conditions (B) and (C) guarantee that the set of nonnegative vectors  $\mathbf{s} = \langle s_1, \dots, s_n \rangle$  satisfying (8.7.13) is a compact, convex, nonvoid,  $n$ -dimensional bounded set. Let us denote this set by  $S$ . Let  $\mathbf{c}$  be an interior point of  $S$ , and let  $S_\epsilon$  be the set of all  $\mathbf{s}$  in  $S$  which satisfy the property that  $\mathbf{s} = \epsilon \mathbf{c} + (1 - \epsilon) \mathbf{s}'$  for  $\mathbf{s}'$  in  $S$  and  $0 < \epsilon < 1$ . The set  $S_\epsilon$  is an interior subset such that the boundary of  $S_\epsilon$  is traversed as  $\mathbf{s}'$  traverses the boundary of  $S$ . By the projection  $\mathbf{P}_\epsilon$  of  $S$  into  $S_\epsilon$  we shall mean

$$\mathbf{P}_\epsilon \mathbf{s} = \epsilon \mathbf{c} + (1 - \epsilon) \mathbf{s}.$$

Our objective is to establish the existence of nonnegative solutions  $\langle p_i \rangle$

and  $\langle s_j \rangle$  to the system of equations (8.7.12) and (8.7.13), with the added feature that

$$\sum_{i=1}^m p_i \left( r_i - \sum_{j=1}^n a_{ij} s_j \right) = 0.$$

The analysis depends upon the use of the Kakutani fixed-point theorem and the duality theorem of linear programming. For every  $s_0$  in  $S_\epsilon$ , we construct the vector  $\sigma = \langle \sigma_j \rangle = \langle f_j(s_0) \rangle$ . Consider the following two problems:

I. For  $\mathbf{A}\mathbf{u} \leq \mathbf{r}$ , where  $\mathbf{u} \geq \mathbf{0}$ , maximize  $(\sigma, \mathbf{u})$ .

II. For  $\mathbf{p}\mathbf{A} \geq \sigma$ , where  $\mathbf{p} \geq \mathbf{0}$ , minimize  $(\mathbf{r}, \mathbf{p})$ .

Since the set of allowable vectors  $\mathbf{u}$  traverses  $S$ , a bounded, closed, convex subset of the first orthant, a solution to Problem I exists. Let the set of all solutions be denoted by  $\bar{T}(s_0)$ . The projection  $\mathbf{P}_\epsilon[\bar{T}(s_0)] = T(s_0)$  is closed, nonvoid, and convex. The fundamental duality theorem (Theorem 5.4.1) tells us that there exists a solution  $\mathbf{p}^0$  to Problem II which has the property that  $(\mathbf{p}^0, \mathbf{r} - \mathbf{A}\mathbf{u}^0) = 0$  for any  $\mathbf{u}^0$  in  $\bar{T}(s_0)$ . The point-to-set mapping  $s_0 \rightarrow T(s_0)$  fulfills all the conditions needed to apply the Kakutani fixed-point theorem. Consequently, there exists an  $s_\epsilon^*$  in  $S_\epsilon$  such that  $s_\epsilon^*$  is also in  $T(s_\epsilon^*)$ .

Let  $s^*$  denote a limit point of  $s_\epsilon^*$  with the maximal number of zero coordinates. We first prove that no coordinate of  $s^* = \langle s_1^*, s_2^*, \dots, s_n^* \rangle$  can be zero. Suppose that the indices  $I_0$  have  $s_{i_0}^* = 0$  for  $i_0$  in  $I_0$ ; then the functions  $f_j(s_\epsilon^*)$  tend to  $\infty$  as  $\epsilon \rightarrow 0$  for every  $j$  in  $I_0$ , but remain bounded in the other coordinates. For  $\epsilon$  sufficiently small the vector  $s_\epsilon^*$  is close to the set of  $u$  which maximizes  $(f, u)$ , where  $u$  is subject to  $\mathbf{A}u \leq \mathbf{r}$  and  $r$  is strictly positive. It follows, however, that some of the coordinates of  $s_\epsilon^*$  with indices in  $I_0$  must be chosen bounded away from zero; for if the contrary were true and all the coordinates of  $s_\epsilon^*$  with  $i$  in  $I_0$  converged to zero, then a larger value of  $(f(s_\epsilon^*), u)$  could be obtained by choosing a  $u$  in  $S$  with the  $i_0$  coordinate strictly greater than a fixed positive  $\epsilon_0$  for every  $i_0$  in  $I_0$ , and the set  $T(s_\epsilon^*)$  could not contain  $s_\epsilon^*$  as constructed. Consequently, in order to avoid these absurdities, we must have every coordinate of  $s^*$  strictly bounded away from zero, and the corresponding  $f_j(s^*)$  are therefore uniformly bounded.

Since the associated solutions  $\mathbf{p}_\epsilon^0$  to Problem II for the particular choice of  $\sigma$  corresponding to  $s_0 = s_\epsilon^*$  minimize  $(\mathbf{r}, \mathbf{p})$  and each  $r_j$  is strictly positive, it follows that the  $\mathbf{p}_\epsilon^0$  lie in a bounded portion of the first quadrant, and hence a limit point  $\mathbf{p}^0$  exists as  $\epsilon$  tends to zero. The vectors  $s^*$  and  $\mathbf{p}^0$  satisfy the following relations:

$$r_i \geq \sum_{j=1}^n a_{ij} s_j^* \quad (i = 1, \dots, m). \quad (8.7.14)$$

$$\sigma_j \leq \sum_{i=1}^m a_{ij} p_i^0 \quad (j = 1, \dots, n), \quad (8.7.15)$$

$$\sigma_j = f_j(s_1^*, \dots, s_n^*) \quad (j = 1, \dots, n). \quad (8.7.16)$$

The linear programming analysis shows that  $\mathbf{s}^*$  solves Problem I, with  $\sigma_j = f_j(\mathbf{s}^*)$ , and  $\mathbf{p}^0$  solves Problem II. Since each coordinate of  $\mathbf{s}^*$  is positive, equality must hold in (8.7.15) (by Theorem 5.4.3). Again, by virtue of the same theorem, we obtain

$$\sum_{i=1}^m p_i^0 \left( r_i - \sum_{j=1}^n a_{ij} s_j^* \right) = 0. \quad (8.7.17)$$

The following theorem summarizes the preceding discussion.

► **THEOREM 8.7.4.** If  $r_i$ ,  $a_{ij}$ , and  $f_j$  satisfy conditions (A), (B), and (C), then there exist two positive vectors  $\mathbf{s}^*$  and  $\mathbf{p}^0$ , the first strictly positive, satisfying (8.7.14), (8.7.16), and (8.7.17), with the equality sign for every equation of the set (8.7.15).

**\*8.8 The Arrow-Debreu equilibrium model of a competitive economy.** We suppose that there are a finite number  $r$  of distinct commodities (including services of all kinds) available. If the same commodity is bought and sold at  $m$  different locations or at  $m$  different times, it may be regarded as  $m$  distinct commodities.

The commodities are produced in production units (e.g., firms). We assume that for each production unit  $j$  ( $j = 1, 2, \dots, n$ ) there is a set  $Y_j$  of possible production plans. An element  $\mathbf{y}_j$  of  $Y_j$  is a vector in  $E^r$  whose  $h$ th component  $y_{hj}$  designates the output of commodity  $h$  according to that plan. As usual, inputs are treated as negative components.  $Y = \sum_j Y_j$  represents the set of all possible input-output schedules of the whole production sector. We make the following assumptions:

I(a).  $Y_j$  is a compact subset of  $E^r$  containing 0.

I(b).  $Y$  is convex.

The convexity assumption reflects the principle of nonincreasing marginal rates of transformation. Moreover, we assume the existence of a finite number of consumption units (individuals or institutions).

Let  $X_i$  ( $i = 1, \dots, m$ ) represent the set of consumption vectors in  $E^r$  available to consumer  $i$ . We now make another assumption:

II.  $X_i$  is a closed convex subset of  $E^r$ , with the additional property that whenever  $\mathbf{x}_i^{(n)} \in X_i$  and some coordinate  $x_{h_0 i}^n$  of  $\mathbf{x}_i^n$  tends to  $\infty$ , then  $x_{h i}^n \rightarrow \infty$  for all coordinates  $h = 1, 2, \dots, r$ .

We shall suppose that there is defined a utility index function  $U_i(\mathbf{x}_i)$  for the  $i$ th consumer such that:

III(a).  $U_i(\mathbf{x}_i)$  is a continuous differentiable function on  $X_i$ .

III(b). If  $U_i(\mathbf{x}_i) \geq U_i(\mathbf{x}'_i)$  and  $\mathbf{x}_i, \mathbf{x}'_i \in X_i$  ( $\mathbf{x}_i \neq \mathbf{x}'_i; 0 < t < 1$ ), then  $U_i[t\mathbf{x}_i + (1-t)\mathbf{x}'_i] > U_i(\mathbf{x}'_i)$ .

III(c). For any  $\mathbf{x}_i$  in  $X_i$ , there is an  $\mathbf{x}'_i$  in  $X_i$  such that  $U_i(\mathbf{x}'_i) > U_i(\mathbf{x}_i)$ .

Assumption III(b) guarantees that the set  $\{\mathbf{x} | U_i(\mathbf{x}) \geq c\}$  is a strictly convex set [see Section B.4, property (ii)] for each constant  $c$ . The function  $U$  clearly induces a preference relation as defined in Section 8.6. Without any budgetary restrictions the  $i$ th consumer would make his selection from  $X_i$  strictly on the basis of his preference scale.

Assumption III(c) states that if we disregard budget considerations the desires of consumers are never completely satisfied.

Let  $\xi_i$  represent the commodities which the  $i$ th consumer holds for initial trading. We assume that:

IV. There exists an  $\mathbf{x}_i$  in  $X_i$  such that  $\xi_i \gg \mathbf{x}_i$ .

Although assumption IV is too strong and not altogether realistic, unfortunately it cannot be completely eliminated. It can be weakened slightly, but only at the expense of making the arguments more tedious and no more rewarding. (For extensions see the notes at the close of the chapter.)

Finally, we assume that the  $i$ th consumer has a contractual claim to the share  $\alpha_{ij}$  of the profit of the  $j$ th production of each commodity such that

V.  $\alpha_{ij} \geq 0$  for all  $i, j$ , and

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

for all  $j$ .

Let  $\mathbf{p} = \langle p_1, \dots, p_r \rangle$  represent a vector of prices of the commodities. Since only relative prices are important, we may normalize  $\mathbf{p}$  as follows without restricting generality:

$$p_v \geq 0, \quad \sum_{v=1}^r p_v = 1.$$

*Definition of a competitive equilibrium.* The existence of a competitive equilibrium is equivalent to the existence of a set of vectors

$$\{\mathbf{p}^*; \mathbf{x}_1^*, \dots, \mathbf{x}_m^*; \mathbf{y}_1^*, \dots, \mathbf{y}_n^*\}$$

such that

$$(A) \quad \mathbf{p}^* \in P = \left\{ \mathbf{p} | \mathbf{p} \in E^r, \quad \mathbf{p} \geq \mathbf{0}, \quad \sum_{v=1}^r p_v = 1 \right\};$$

(B) for each  $j$

$$(\mathbf{p}^*, \mathbf{y}_j^*) = \max_{y_j \in Y_j} (\mathbf{p}^*, \mathbf{y}_j);$$

(C) for each  $i$ ,  $\mathbf{x}_i^*$  maximizes  $U_i(\mathbf{x}_i)$  over the set

$$\left\{ \mathbf{x}_i \mid \mathbf{x}_i \in X_i, \quad (\mathbf{p}^*, \mathbf{x}_i) \leq (\mathbf{p}^*, \xi_i) + \sum_{j=1}^n \alpha_{ij}(\mathbf{p}^*, \mathbf{y}_j^*) \right\};$$

(D)  $\mathbf{x}^* \leq \mathbf{y}^* + \xi^*$  and  $(\mathbf{p}^*, \mathbf{x}^* - \mathbf{y}^* - \xi^*) = 0$ ,

where

$$\mathbf{x}^* = \sum_{i=1}^m \mathbf{x}_i^*, \quad \mathbf{y}^* = \sum_{j=1}^n \mathbf{y}_j^*.$$

Condition (B) asserts that production is such as to maximize profits when the prices are the equilibrium market prices. Condition (C) states that the  $i$ th consumer seeks to maximize his utility with respect to all available consumption vectors, where a consumption vector is possible if it belongs to his availability set  $X_i$  and if its cost does not exceed his income. His income derives from two sources: first, the income associated with his initial holdings; and second, the income derived from his share of the profits of all production. In other words, condition (C) expresses the budget restraint under conditions of equilibrium. Condition (D) states that all demand is fulfilled, and further that any commodity supplied in excess of demand is priced zero.

The basic theorem is as follows:

► **THEOREM 8.8.1.** For any system satisfying assumptions I–V, a competitive equilibrium exists.

The proof of this theorem is a simple consequence of the following topological result concerning point-to-set mapping, which also is of some independent interest.

► **LEMMA 8.8.1.** Let  $S$  be a bounded, upper semicontinuous, set-valued function mapping the set  $P$  into  $E^r$  such that (a)  $S(\mathbf{p})$  is a nonvoid convex set for all  $\mathbf{p} \in P$ , and (b) if  $\mathbf{z} \in S(\mathbf{p})$ , then  $(\mathbf{p}, \mathbf{z}) \geq 0$ . Then there exist values  $\mathbf{p}$  in  $P$  and  $\mathbf{z}$  in  $S(\mathbf{p})$  such that  $\mathbf{z} \geq 0$ .

(The mapping  $S(\mathbf{p})$  as upper semicontinuous is to be understood in the usual sense, i.e., if  $\mathbf{z}_n \in S(\mathbf{p}_n)$  and  $\mathbf{z}_n \rightarrow \mathbf{z}_0, \mathbf{p}_n \rightarrow \mathbf{p}_0$ , then  $\mathbf{z}_0 \in S(\mathbf{p}_0)$ ; see also Section C.2.)

*Proof.* Suppose that the lemma is false. This means that each  $S(\mathbf{p})$  does not touch the nonnegative orthant. In that case we claim that there exists a positive  $\epsilon_0$  such that the family of convex sets

$$T(\mathbf{p}) = \{ \mathbf{w} \mid \mathbf{w} = \mathbf{z} + \epsilon_0 \mathbf{p}, \quad \mathbf{z} \in S(\mathbf{p}) \}$$

for each  $\mathbf{p}$  in  $P$  also does not touch the nonnegative orthant. Otherwise,

there must exist a sequence  $\mathbf{p}_n \rightarrow \tilde{\mathbf{p}}$  and  $\mathbf{z}_n \in S(\mathbf{p}_n) \rightarrow \mathbf{z}$  such that  $\mathbf{z}_n + \epsilon_n \mathbf{p}_n \geq 0$ , with  $\epsilon_n \rightarrow 0$  ( $\mathbf{z}_n$  may be chosen to converge, since all  $S(\mathbf{p}_n)$  are compact and lie in a bounded region); and it follows from the continuity of  $S$  that  $\mathbf{z}$  is in  $S(\tilde{\mathbf{p}})$  and  $\mathbf{z} \geq 0$ , in contradiction to our assumption. Hence  $T(\mathbf{p})$  is disjoint from the nonnegative orthant as stated.

We notice next that a plane separating the positive orthant and one of the sets  $T(\mathbf{p})$  may be identified with a vector  $\mathbf{q} = \langle q_1, \dots, q_r \rangle$  such that

$$q_i \geq 0, \quad \sum_{i=1}^r q_i = 1, \quad (\mathbf{q}, \mathbf{w}) \leq 0$$

for  $\mathbf{w}$  in  $T(\mathbf{p})$ , and conversely (see Lemma B.1.3).

We now construct a point-to-set mapping as follows: For each  $\mathbf{p}$  in  $P$  let  $Q(\mathbf{p}) \subset P$  consist of all normalized vectors in  $P$  representing all separating planes of the positive orthant and  $T(\mathbf{p})$ . Since  $T(\mathbf{p})$  does not touch the positive orthant, it follows that  $Q(\mathbf{p})$  is nonvoid (Lemma B.1.3). Clearly,  $Q(\mathbf{p})$  is convex and closed and describes an upper semicontinuous point-to-set mapping; the nature of the continuity of  $Q(\mathbf{p})$  follows easily from the corresponding continuity of  $T(\mathbf{p})$ . By the Kakutani theorem (see Section C.2) there exists  $\mathbf{p}_0 \in Q(\mathbf{p}_0)$ . But for this  $\mathbf{p}_0$ , it follows from hypothesis (b) that  $(\mathbf{p}_0, \mathbf{z}) \geq 0$  when  $\mathbf{z}$  is in  $S(\mathbf{p}_0)$ , and hence  $(\mathbf{p}_0, \mathbf{w}) > 0$  for  $\mathbf{w}$  in  $T(\mathbf{p}_0)$ . This inference is incompatible with the meaning of  $\mathbf{p}_0$  in  $Q(\mathbf{p}_0)$ . Hence our original supposition was false and the lemma is proved.

*Proof of Theorem 8.8.1.* For each  $\mathbf{p}$  in  $P$ , we construct  $S(\mathbf{p})$  with the properties of Lemma 8.8.1 as follows. First define  $Y_j(\mathbf{p})$  as the set of all vectors in  $Y_j$  which maximize  $(\mathbf{p}, \mathbf{y}_j)$ . Let  $Y(\mathbf{p}) = \sum Y_j(\mathbf{p})$ , and let

$$\beta_j(\mathbf{p}) = \max_{u_j \in Y_j} (\mathbf{p}, \mathbf{y}_j).$$

It is easily established that  $Y(\mathbf{p})$  is a closed, nonvoid, convex set and defines a bounded upper semicontinuous point-to-set mapping. Next let  $X_i(\mathbf{p})$  designate the set of all  $\mathbf{x}_i$  in  $X_i$  which maximize  $U_i(\mathbf{x}_i)$  subject to the condition

$$(\mathbf{p}, \mathbf{x}_i) \leq (\mathbf{p}, \xi_i) + \sum_{j=1}^n \alpha_{ij} \beta_j(\mathbf{p}). \quad (8.8.1)$$

An important observation, which we shall call on later, is that for  $\mathbf{x}_i$  in  $X_i(\mathbf{p})$  equality must prevail in (8.8.1). Let us suppose the contrary. Then by assumption III(c) there exists an  $\mathbf{x}'_i \in X_i$  such that  $U_i(\mathbf{x}'_i) > U_i(\mathbf{x}_i)$ , and by assumptions III(a) and III(b) we can find  $\tilde{\mathbf{x}}_i = t\mathbf{x}_i + (1-t)\mathbf{x}'_i$  for appropriate  $t$  ( $0 < t < 1$ ) such that  $U_i(\tilde{\mathbf{x}}_i) > U_i(\mathbf{x}_i)$  and  $\tilde{\mathbf{x}}_i$  still obeys the constraints of (8.8.1), thus contradicting the meaning of  $\mathbf{x}_i$  in  $X_i(\mathbf{p})$ . Hence, equality in (8.8.1) holds for such  $\mathbf{x}_i$ .

From the fact that  $\beta_j(\mathbf{p}) \geq 0$  [recall that  $\mathbf{0} \in Y_i$  by assumption I(a)] and from assumptions I, II, III, and IV, we see immediately that  $X_i(\mathbf{p})$  is a nonvoid, convex, closed, bounded set. Define

$$X(\mathbf{p}) = \sum_{i=1}^n X_i(\mathbf{p}),$$

which is also nonvoid, convex, closed, and bounded. Again, in view of assumptions III and IV, it is readily verified that  $X(\mathbf{p})$  is an upper semicontinuous point-to-set mapping. Finally, let us form

$$S(\mathbf{p}) = \{\mathbf{z} | \mathbf{z} = \mathbf{y} - \mathbf{x} + \boldsymbol{\xi}, \quad \mathbf{y} \text{ in } Y(\mathbf{p}), \quad \mathbf{x} \text{ in } X(\mathbf{p})\},$$

where

$$\boldsymbol{\xi} = \sum_{i=1}^m \xi_i.$$

We now prove that  $S(\mathbf{p})$  is a bounded upper semicontinuous point-to-set mapping from  $P$  to  $E^r$  such that (a)  $S(\mathbf{p})$  is convex, nonvoid, and closed, and (b)  $(\mathbf{p}, \mathbf{z}) = 0$  for each  $\mathbf{z}$  in  $S(\mathbf{p})$ .

We have already noted the truth of (a). To check (b), we see that the terms of the sum of  $\mathbf{y}$  in  $Y(\mathbf{p})$  and  $\mathbf{x}$  in  $X(\mathbf{p})$  obey the relation (8.8.1) with equality, i.e.,

$$(\mathbf{p}, \mathbf{x}_i) = (\mathbf{p}, \xi_i) + \sum_{j=1}^n \alpha_{ij} \beta_j(\mathbf{p}) = (\mathbf{p}, \xi_i) + \sum_{j=1}^n \alpha_{ij}(\mathbf{p}, \mathbf{y}_j),$$

where

$$\mathbf{x} = \sum_{i=1}^m \mathbf{x}_i, \quad \mathbf{y} = \sum_{j=1}^n \mathbf{y}_j.$$

If we sum on  $i$ , it follows from assumption V that

$$(\mathbf{p}, \mathbf{x}) = (\mathbf{p}, \boldsymbol{\xi}) + (\mathbf{p}, \mathbf{y}),$$

which agrees with the statement of (b).

The preparations for invoking the result of Lemma 8.8.1 are here, and we deduce the existence of  $\mathbf{p}^*$  and  $\mathbf{z}^* = \mathbf{y}^* - \mathbf{x}^* + \boldsymbol{\xi}$  such that  $\mathbf{z}^*$  belongs to  $S(\mathbf{p}^*)$  and  $\mathbf{z}^* \geq \mathbf{0}$ . The vector system

$$\{\mathbf{p}^*; \mathbf{x}_1^*, \dots, \mathbf{x}_m^*; \mathbf{y}_1^*, \dots, \mathbf{y}_n^*\}$$

constitutes a competitive equilibrium where

$$\mathbf{x}^* = \sum_{i=1}^m \mathbf{x}_i^*, \quad \mathbf{y}^* = \sum_{j=1}^n \mathbf{y}_j^*.$$

Conditions (A), (B), and (C) (see p. 281) are satisfied by the definition of  $\mathbf{x}_i^*$ ,  $\mathbf{y}_j^*$  and the fact that  $\mathbf{z}^*$  is in  $S(\mathbf{p}^*)$ . Condition (D) follows from the result  $\mathbf{z}^* \geq \mathbf{0}$  and relation (b) above. This completes the proof of the theorem.

**8.9 Problems.** Problems 1–6 embody some simple formulations of macro-economic relations.

1. Suppose that the supply  $S(t)$  and the demand  $D(t)$  for a single commodity are related to its price  $p(t)$  by the expressions

$$S(t) = \alpha + a \left( p(t) + \frac{dp}{dt} \right), \quad D(t) = \beta + bp(t),$$

where  $\alpha$ ,  $a$ ,  $\beta$ , and  $b$  are constants. Assume that in each instance the price is set so that  $S(t) = D(t)$ . Discuss the limiting properties of  $p(t)$ .

2. Let  $Q_t$  denote the stocks on hand at the end of period  $t$ . Suppose that supply and demand are known functions of price at time  $t$  of the form

$$D(t) = \beta + bP(t), \quad S(t) = \alpha + aP(t).$$

Consider a model in which the price varies with the change in stocks according to the law

$$P(t) = P(t-1) - \lambda(Q_{t-1} - Q_{t-2}) \quad (t \geq 2). \quad (*)$$

Show that  $P_t = \bar{P} + [P(0) - \bar{P}]C^t$ , where

$$\bar{P} = \frac{\alpha - \beta}{b - a}, \quad C = 1 - \lambda(b - a).$$

3. Consider the model of Problem 2, taking time to be continuous. The law (\*) becomes

$$\frac{dP}{dt} = -\lambda \frac{dQ}{dt},$$

where

$$Q(t) = Q(0) + \int_0^t [S(\tau) - D(\tau)] d\tau,$$

$D(t) = \beta + bP(t)$ , and  $S(t) = \alpha + aP(t)$ . Determine  $P(t)$  explicitly.

4. A well-known macro-economic relation is formulated as follows: national income = consumption + investment (savings) + autonomous investment, or

$$Y(t) = C(t) + I(t) + A(t).$$

Consider this relation at time  $t$  if  $C(t) = cY(t)$  [ $0 < c < 1$ ; here  $1/(1 - c)$

is called the multiplier],

$$I(t) = v \frac{dY}{dt}$$

[ $v > 0$  is called the accelerator], and  $A = A_0 e^{rt}$ . Determine  $Y(t)$  for all  $t$  and interpret the solution.

5. In the macro-economic model of Problem 4, let

$$C(t) = \gamma Y + c_1 \frac{dY}{dt} + c_2 \frac{d^2Y}{dt^2}, \quad I_t = v \frac{dY}{dt}, \quad A = 0.$$

Determine the form of  $Y(t)$  for all  $t$  and interpret the solution.

6. Consider the model of Problem 4 if time is taken to be discrete. Let  $C(t) = c_1 Y_{t-1} + c_2 Y_{t-2}$ ,  $I(t) = v(Y_{t-1} - Y_{t-2})$ ; and  $A(t) = A = \text{constant}$  for all  $t$ . Determine  $Y(t)$  for all  $t$ .

7. The problem of simple monopoly can be stated as follows: Suppose that it costs  $\phi(\mathbf{x})$  to produce a commodity bundle  $\mathbf{x}$ , and that when  $\mathbf{x}$  is offered on the market the price vector obtained per unit is  $\mathbf{p} = h(\mathbf{x})$ . The net profit on  $\mathbf{x}$  is therefore

$$(\mathbf{p}, \mathbf{x}) - \phi(\mathbf{x}). \quad (**)$$

Determine a relationship satisfied by  $\mathbf{x}$  which maximizes (\*\*).

8. If  $\mathbf{C}$  is a real or complex  $n \times n$  matrix such that

$$S_j = \sum_{i=1}^n |C_{ij}| \leq 1 \quad (j = 1, \dots, n),$$

prove that

$$\prod_{j=1}^n \left(1 - \sum_{i=1}^n |C_{ij}|\right) \leq \det(\mathbf{I} - \mathbf{C}).$$

9. Let  $\mathbf{A}$  be a matrix  $\|a_{ij}\|$  ( $i, j = 1, \dots, n$ ) such that

$$2 \det \begin{pmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{pmatrix} > \sum_{j < k} \left| \det \begin{pmatrix} a_{1j}, a_{1k} \\ a_{2j}, a_{2k} \end{pmatrix} \right|$$

and

$$2|a_{ii}| > \sum_j |a_{ij}| \quad (i = 3, 4, \dots, n).$$

Prove that  $\det \mathbf{A} \neq 0$ .

10. Use Corollary 8.2.1 to show that if  $\mathbf{A}$  is a strictly positive matrix of order  $n$  and has more than one real characteristic root, then all the maximal characteristic numbers of the principal minors of  $\mathbf{A}$  of order  $n - 1$  lie between the two greatest real characteristic numbers of the matrix  $\mathbf{A}$ .

(The maximal characteristic number of a positive matrix is equal to  $\lambda_0(\mathbf{A})$ ; see p. 252.)

11. Show that for the positive number  $\lambda$  to be an upper limit to the moduli of the characteristic numbers of the matrix  $\mathbf{A}$  ( $\mathbf{A} \geq \mathbf{0}$ ), it is necessary and sufficient that the first set of principal minors of the determinant  $\mathbf{B}(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})_1^n$ , that is,

$$b_{11}, \quad \begin{vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{vmatrix}, \quad \begin{vmatrix} b_{11}, & b_{12}, & b_{13} \\ b_{21}, & b_{22}, & b_{23} \\ b_{31}, & b_{32}, & b_{33} \end{vmatrix} \cdots \begin{vmatrix} b_{11}, \dots, b_{1k} \\ b_{21}, \dots, b_{2k} \\ \vdots \\ b_{k1}, \dots, b_{kk} \end{vmatrix} \cdots,$$

be positive.

12. Consider a matrix  $\mathbf{A}$  ( $\mathbf{A} \gg \mathbf{0}$ ) having more than one real characteristic root. Show that between the two greatest real characteristic numbers of  $\mathbf{A}$  there exists at least one principal minor of  $\mathbf{A}$  of order  $n - 2$  which has its maximal characteristic root lying in this interval. (Hint: Use Corollary 8.2.1.)

13. Suppose that the cofactors of a certain row or column in the determinant of  $\mathbf{B}(\lambda) = (\lambda\mathbf{E} - \mathbf{A})$  ( $\mathbf{A} \gg \mathbf{0}, \lambda > 0$ ) are all nonnegative but not all zero. Prove that  $\lambda > \lambda_0(\mathbf{A})$  whenever  $\det \mathbf{B}(\lambda) > 0$ .

14. Let  $\mathbf{A}$  be a positive matrix, and for fixed  $\lambda$  let all algebraic cofactors  $B_{ik}$  of the determinant  $B(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})_1^n$  be positive. Show that the inequalities  $0 < \tilde{a}_{ij} \leq a_{ij}$  ( $i, j = 1, \dots, n$ ) imply the inequalities

$$\det(\lambda\mathbf{I} - \tilde{\mathbf{A}})_1^k \geq \det(\lambda\mathbf{I} - \mathbf{A})_1^k \quad (k = 1, \dots, n).$$

When does equality hold?

15. Let  $\mathbf{B} = \|b_{ij}\|$  be an  $r \times r$  matrix with  $b_{ij} \geq 0$  for  $i \neq j$ . Suppose that there exists  $\mathbf{z}$  such that  $\mathbf{z} < \mathbf{0}, \mathbf{Bz} \gg \mathbf{0}$ . Show that  $\mathbf{B}$  is stable. (A matrix is called stable if all its eigenvalues have negative real parts.)

16. A product  $y$  is produced by the use of two factors  $x_1$  and  $x_2$  according to the linear homogeneous production function  $y = kx_1x_2$ , where  $k$  is a constant. Given the factor prices  $p_1$  and  $p_2$ , find the optimum usage of the factors yielding a minimum cost for the product  $y$ . Show that at this optimum  $x_1 = c_1y$  and  $x_2 = c_2y$ .

17. Prove that for any matrix  $\mathbf{C}$  there exist nonnegative vectors  $\mathbf{x}$  and  $\mathbf{y}$  such that

$$\begin{aligned} \mathbf{x}\mathbf{C} &\geq \mathbf{0}, & \mathbf{C}\mathbf{y} &\leq \mathbf{0}; \\ \mathbf{x} - \mathbf{C}\mathbf{y} &\gg \mathbf{0}, & \mathbf{y} + \mathbf{x}\mathbf{C} &\gg \mathbf{0}. \end{aligned}$$

18. Using the Arrow-Debreu model and assuming that the technology is subject to constant returns to scale (i.e., each  $Y_j$  is a closed convex cone),

show that in an equilibrium the profit of each firm is zero—and hence that an equilibrium is determined without regard to  $\alpha_{ij}$  (see p. 281).

19. (a) Prove that for  $n = 2$  the strong axiom of revealed preference is implied by the weak axiom.

(b) Prove that if the demand vector function  $\mathbf{f}(\mathbf{p}, M)$  satisfies the weak axiom of revealed preference, then  $\mathbf{f}(\mathbf{p}, M)$  is positively homogeneous of order 0, that is, that  $\mathbf{f}(\lambda\mathbf{p}, \lambda M) = \mathbf{f}(\mathbf{p}, M)$  for any  $\lambda > 0$ .

#### NOTES AND REFERENCES TO CHAPTER 8

It is generally agreed that the mathematical method began to play a significant role in economics with A. A. Cournot's *Recherches sur les principes mathématiques de la théorie des richesses*, which was published in 1838, although a few economists like Ceva and von Thunen contributed to the subject before Cournot. In the 1870's W. Stanley Jevons, Léon Walras, and Karl Menger created the subjective theory of value, which distinguishes the modern economics from the classical. Of the three, Walras, who was greatly influenced by Cournot, was most prominent; his theory of general equilibrium did for economics what Newton did for physics. Among those who may be classified as belonging to the Lausanne school founded by Walras are Pareto, Barone, Pantaleoni, Schumpeter, Leontief, and Hicks. Menger himself lacked the appropriate mathematical tools, but his marginal theory of value and much of the other work of the Austrian school of economics, which he helped to found, had a decisive influence on modern mathematical economics.

In England F. Y. Edgeworth's *Mathematical Psychics* (1881) proposed a detailed treatment of exchange. Some years later Alfred Marshall created the Cambridge school of economic analysis, based on a proper balance of intuitive knowledge and scientific precision; prominent later members of this school were Alphonse Pigou, John Maynard Keynes, and D. H. Robertson. It is accurate to say that the majority of contemporary research in economics derives ultimately from the massive contributions of the Lausanne, Austrian, and Cambridge schools.

The two main sources of current developments in mathematical economic theory are J. R. Hicks, *Value and Capital* [121], and P. A. Samuelson, *Foundations of Economic Analysis* [210], which was written at Harvard as a dissertation in 1937 and published in book form later.

Two very useful textbooks, both covering a multitude of topics of mathematical economics, are Allen [2] and Dorfman, Samuelson, and Solow [65].

The subject of mathematical economics divides into two main parts, macro-economic theory and micro-economic theory. The first is concerned chiefly with relations of economic variables aggregated over the system, the second with relations between individual economic variables, notably consumption units, production units, and market prices of commodity bundles.

Mathematical economics may also be divided into static theory and dynamic theory. A static problem is a problem in which time does not explicitly appear;

a dynamic problem is one in which the time variable is directly involved. Of course, even when we are confronted with a dynamic problem, it may happen that the solution is independent of time. We have then a stationary phenomenon. Frequently, but not always, a static situation may be interpreted as a stationary state of a dynamic system.

Research work in micro-economic theory has been concerned mainly with statics, while the majority of research work in macro-economic theory has been in dynamics. A typical procedure in macro theory is to propose models of aggregated consumption, investment, and national income as related functions of time, and then analyze the convergence or oscillatory character of, say, the income function over time.

We have not mentioned the well-developed statistical point of view of econometrics. From this point of view the statistical problem is to estimate the parameters underlying the dynamic system by methods embracing modifications of regression analysis. Econometric problems have been studied by J. Koavelma, T. Koopmans, A. Wald, T. W. Anderson, H. Rubin, L. Hurwicz, J. Marschak, N. Schultz, I. Timberger, and H. Wold. The leading textbook is L. R. Klein, *A Textbook of Econometrics*.

**§ 8.1.** The first linear model of equilibrium and exchange was proposed by Walras, who developed a system of equations to describe the simple phenomenon of equilibrium. However, he did not pursue the mathematical questions of the existence and uniqueness of solutions to these equations. With the aid of the theory of positive matrices this gap has been filled adequately by many workers in the field.

Until about 1948, the study of equilibrium models was largely restricted to the linear case.

The models of this section follow the formulation of open and closed systems due to Leontief (see [166] and Morgenstern [182]).

**§ 8.2.** The theory of matrices with nonnegative entries begins with Frobenius [89]. Since 1908 there have been innumerable extensions and applications of the Frobenius results. The generalizations have followed two main directions: (1) replacing the property of positivity by the requirement that the matrix  $A$  keep a cone invariant, and (2) pursuing the infinite-dimensional analogs of the Frobenius theorems involving positive kernels. An impressive collection of these extensions is found in Gantmacher and Krein [102], which in addition develops much of the finer structure of the theory of positive matrices. For abstract treatments of this subject see Krein and Rutman [153] and Karlin [139]. Unfortunately, economists continue to rediscover many of these theorems and to assign them thoroughly inaccurate priorities.

Direct interpretations of the theory of positive matrices with reference to economic theory, and various expositions of the properties of positive matrices and  $M$ -matrices, have been presented by Metzler [180], Mosak [185], Debreu and Herstein [63], Solow [227], Hawkins and Simon [119], and many others.

The simple proofs of this section represent a direct paraphrase of the abstract arguments of Krein and Rutman [153] and Karlin [139]. The characterization of the largest eigenvalue according to Corollary 8.2.1 seems to be new and appears

very useful, as many of the later arguments indicate. The elegant argument proving part (d) of Theorem 8.2.3 is new and due to Bohnenblust (private communication).

Results related to Theorem 8.2.4 can be found in Kotelyanskii [151].

For a discussion of the classification of a matrix into its periodic and non-periodic components, see Feller [80, Chaps. 13 and 15].

The fact that results valid for  $M$ -matrices can be translated to positive matrices and vice versa is well known. The statement embodying (8.3.4) is known to economists as the Hawkins-Simon criterion [119] (see also Kotelyanskii [151]).

**§ 8.3.** For other results of the kind developed in this section, see Morgenstern [182].

**§ 8.4.** The modern development of the theory of production is attributed to Walras and Marshall. Samuelson devotes Chap. iv of his *Foundations* to production theory, see also Hicks [121] and Dorfman [64].

The current theory of production is formulated in terms of activity analysis models, utilizing the theorems of linear and nonlinear programming. As general references on activity analysis, we cite the compilations edited by Koopmans [146] and Morgenstern [182]. The nonlinear programming paper of Kuhn and Tucker [164] is basic to the finer mathematical analysis of production theory.

Inseparable from the theory of production is the theory of the distribution of wealth (imputed prices). Information on the distribution of wealth is obtained by solving the dual problem of the associated programming problem and suitably interpreting the concept of shadow price (see Uzawa [234]).

**§ 8.5.** Theorem 8.5.1 was suggested by Samuelson. Koopmans [148] proved the theorem for the case of three commodities, and Arrow [3] established the result in general. An alternative proof was advanced by Gale [94]. The proof given here is new.

**§ 8.6.** The modern theory of consumption originated with Walras and Jevons. Slutsky [225] and later Hicks [121] greatly elaborated this theory. Slutsky's analysis relies on the existence of a utility indicator function, which is used in deducing the Slutsky relations satisfied by the demand functions (see Theorem 8.6.1). McKenzie [177] derived the same conclusions without the use of a utility indicator function. Our proof is a modification of McKenzie's. The concept of the weak principle of revealed preference was introduced by Samuelson [211], who also showed that a demand function induced by an indicator function implies the validity of this principle.

Houthakker [126] proved that the strong axiom of revealed preference is equivalent to the existence of the utility function.

An axiomatic approach to the analysis of preference relations was carried out by Uzawa [232]. He established a one-to-one correspondence between a preference relation induced by the strong axiom of revealed preference and a general preference relation satisfying postulates P-1 through P-5. The discussion of this section is based on Uzawa's paper.

**§ 8.7.** The main advances to date in the study of equilibrium models are associated with the names of Walras, Wald, McKenzie, and Arrow and Debreu. To Walras [252] we owe the first sophisticated formulation of the equilibrium phenomenon (see notes to § 8.1). Wald [251] later carried out a comprehensive mathematical analysis of Walras's equations and established the existence and uniqueness of solutions to these equations under a variety of conditions. Finally, McKenzie, and Arrow and Debreu [7] independently formulated general models of equilibrium and proved the existence of a competitive equilibrium. None of the three models (Walras-Wald, McKenzie, Arrow-Debreu) subsumes any of the others as a special case; rather, all three should be viewed as variant descriptions of equilibrium in a competitive economy. Alternative analyses of the same models or other versions of equilibrium have been proposed independently by Gale, McKenzie, and Uzawa.

The statements of Theorems 8.7.1–3 were discovered by Gale [93]. The proofs evolved here appear to be more direct.

Our analysis of the Wald equations, which combines the Kakutani theorem and the duality theorem of linear programming, follows the analysis of Kuhn [159].

**§ 8.8.** The setup of the model here is that of Arrow and Debreu, with some small changes. The arguments of this section appear to be new and simpler than the original arguments, although it should be pointed out that the Arrow-Debreu formulation is slightly more general and undoubtedly more realistic.

Lemma 8.8.1 was stated by Gale [93], who appeals to the Knaster-Marecniewicz fixed-point theorem for its proof. Direct application of the Kakutani fixed-point theorem gives a much simpler proof.

## CHAPTER 9

### MATHEMATICAL METHODS IN THE STUDY OF ECONOMIC MODELS (Continued)

In this chapter we examine aspects of welfare economics, the stability of equilibrium price vector systems, and the theory of balanced growth.

Equilibrium models are essentially descriptive rather than normative; their function is simply to describe as accurately as possible the interactions of supply, demand, and prices in a competitive economy. By contrast, the theory of welfare economics is essentially normative; it seeks to determine production and consumption vectors which will yield optimal results from a definite point of view, that of the consumer. Equilibrium theory is concerned with things as they are, welfare theory with things as they might be.

The goal of welfare economics is a system in which all consumers uniformly achieve their maximal utility. When more than one consumer is involved, this rarely happens. In this case we characterize the form of all efficient consumption vectors which are so constituted that the utility of any one consumer cannot be increased without making another consumer worse off. This concept is known as the "Pareto optimum." Certainly, any consumption program should be formulated in terms of a Pareto optimum. However, since there exist a multitude of Pareto optimum consumption vectors, a further criterion must be imposed to single out a specific optimum consumption vector system. One such criterion is noted in Section 9.1.

The mathematics of the theory of welfare economics is in the main a non-trivial application of the vector maximum nonlinear programming problem of Section 7.4. The connections of welfare economics to equilibrium theory are also brought out in Section 9.1.

In the remainder of this chapter we consider various dynamic problems of mathematical economics. In Section 9.2 we formulate several versions of the process by which prices respond to changes from period to period in the excess demand vector functions (i.e., changes in the amount by which total demand exceeds total supply). The excess demand functions are themselves functions of the prices of the previous period. For the equilibrium price vector, the excess demand functions are essentially zero.

When time is considered continuous, the price adjustment process is described by a system of differential equations. The problem of determining whether this process is stable is essentially a matter of determining conditions which cause the price vector as a function of time to converge with increasing time to an equilibrium price vector. Mathematically speaking, the problem is one of determining criteria which ensure the

stability of solutions of a special system of differential equations (or difference equations, depending on the model).

The nature of these criteria depends on economic considerations. For example, the concepts of "gross substitutes" and "gross complements" (see pp. 305 and 307) lead to natural conditions satisfied by the excess demand functions; among other things, they permit us to pair off commodities according to the manner in which a rise in the price of one commodity affects the demand for other commodities. The mathematics of the study of the gross substitutes concept involves some intriguing relations for special vector fields of functions (see in particular Section 9.4). Another meaningful economic concept which implies stability of the price adjustment process is the principle of revealed preference.

In Sections 9.2 through 9.5 we investigate the stability structure of the differential-equations formulation of the price adjustment process for a variety of conditions. In Section 9.6 we examine a difference-equation version of the same process. In Sections 9.7 and 9.8 we explore aspects of the theory of local stability of the adjustment process when the excess demand functions depend not only on current prices but also on the so-called "anticipated future prices" of the process. The two versions of stability (with and without the factor of anticipated future prices) are compared.

Sections 9.9 and 9.10 are concerned with a different kind of dynamic process: the growth of production and consumption in successive periods. One fundamental corollary shows that for a suitable hypothesis the rate of expansion is equal to the rate of interest. In Section 9.9 we describe the classical model of an expanding economy formulated by von Neumann. The final section presents a generalization of this model and an analysis of one of its special cases, the Leontief dynamic model.

**\*9.1 Welfare economics.** A descriptive study of the economy can be characterized mainly by the phenomenon of equilibrium. An equilibrium state is a situation such that with respect to a set of prices no consumer can make himself better off without spending more, and no producer can make a larger profit. A study of welfare economics is mainly concerned with the problem of describing the states of the economy where no consumer can be made better off without making another consumer worse off.

We turn now to a detailed analysis of welfare theory. Commodity bundles are represented by vectors of size  $r$ . To the  $i$ th consumer ( $i = 1, 2, \dots, m$ ) corresponds a set  $X_i$  of consumption vectors in  $E^r$ , and a utility indicator function  $U_i(\mathbf{x}_i)$  defined for  $\mathbf{x}_i$  in  $X_i$  which characterizes

---

\* Starred sections are advanced discussions that may be omitted at first reading without loss of continuity.

the  $i$ th consumer's preference scale. (The preference relation induced by  $U_i(\mathbf{x}_i)$  is assumed to satisfy all the usual properties enumerated in Section 8.6. For convenience of exposition we shall assume further that  $U_i(\mathbf{x}_i)$  defines a strictly concave function for  $\mathbf{x}_i$  in  $X_i$ .) The economy attains perfection when all consumers can simultaneously achieve their maximal utility. We shall see that the equations of a competitive economy generally possess no solution consistent with the ability of all consumers to achieve maximal utility.

A vector system  $\{\mathbf{x}_i\}$  ( $\mathbf{x}_i \in X_i$ ) is said to be a Pareto optimum if (a)  $\{\mathbf{x}_i\}$  is possible (see below) and (b) there exists no other vector system  $\{\mathbf{x}'_i\}$  ( $\mathbf{x}'_i \in X_i$ ) such that

$$U_i\{\mathbf{x}'_i\} \geq U_i(\mathbf{x}_i) \quad (i = 1, \dots, m),$$

with strict inequality for at least one consumer. (This concept is in essence the same as the notion of "admissibility" in statistics.)

Our objective in this section is twofold: first, we seek to characterize all Pareto optimum consumption vector systems; second, we examine the relationship between a given Pareto optimum and the existence of a competitive equilibrium (see Section 8.8).

At the close of the section we discuss briefly the problem of how to select a Pareto optimum vector system from the set of all such efficient vector systems.

We first consider the question of when a Pareto optimum is a competitive equilibrium and conversely.

*The case of a single consumer.* Let  $Y_j$  ( $j = 1, \dots, n$ ) in  $E^r$  designate the set of production possibility vectors of firm  $j$ . Clearly,

$$Y = \sum_{j=1}^n Y_j$$

constitutes the vector of the total production of the economy.

Let  $X$  contained in the positive orthant of  $E^r$  represent the consumption set available to the consumer, and let  $U(\mathbf{x})$  denote his utility index function, which we assume to be a strictly concave function. The total production set  $Y$  is assumed to be convex and compact. The consumer desires to select a vector  $\mathbf{x}^*$  from  $X$  such that

$$U(\mathbf{x}^*) = \max U(\mathbf{x}),$$

where the maximum is extended over all  $\mathbf{x}$  in  $X$  for which there exist vectors  $\mathbf{y}$  in  $Y$  such that  $\mathbf{x} \leq \mathbf{y}$ .

Any vector  $\mathbf{x}$  in  $X$  for which there exists  $\mathbf{y}$  in  $Y$  such that  $\mathbf{y} \geq \mathbf{x}$  is called feasible or possible. In other words, the consumer acts in a manner to maximize his utility, taking account of the technological possibilities.

Since  $Y$  is compact, it is clear that the range of possible  $\mathbf{x}$  vectors ( $\mathbf{x} \leq \mathbf{y}$ ) is also compact, and hence the operation  $\max U(\mathbf{x})$ , extended only over possible consumption vectors, is well defined.

We shall also impose the following requirements:

*Assumption I.* There exist vectors  $\mathbf{x}^0$  in  $X$  and  $\mathbf{y}^0$  in  $Y$  such that  $\mathbf{y}^0 - \mathbf{x}^0 \gg 0$ .

*Assumption II.* For each  $\mathbf{x}$  in  $X$  there exists a vector  $\mathbf{x}'$  in  $X$  such that  $U(\mathbf{x}') > U(\mathbf{x})$ .

Assumption II states that if we disregard budget restrictions and feasibility, there is no consumption vector which is a point of saturation (see p. 281).

The problem of determining the optimum consumption vector can be put into the form of a concave programming problem as follows.

For  $\{\mathbf{x}, \mathbf{y}\}$  consisting of any pair in  $X \otimes Y$ , we define

$$g(\mathbf{x}, \mathbf{y}) = U(\mathbf{x}).$$

Let  $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_2, \dots, \mathbf{y}_r - \mathbf{x}_r \rangle$  denote a vector system of functions defined everywhere in  $E^r \otimes E^r$ . In this notation the problem of selecting a consumption vector of maximal utility becomes

$$\max g(\mathbf{x}, \mathbf{y}) = \max U(\mathbf{x}) \quad (9.1.1)$$

subject to the constraints

$$\mathbf{F}(\mathbf{x}, \mathbf{y}) \geq 0 \quad (\mathbf{x} \in X; \mathbf{y} \in Y). \quad (9.1.2)$$

The definition of a competitive equilibrium in the case of one consumer is as follows: A vector system  $\{\mathbf{p}^*, \mathbf{x}^*, \mathbf{y}^*\}$ , where  $\mathbf{y}^*$  is in  $Y$ ,  $\mathbf{x}^*$  is in  $X$ , and  $\mathbf{p}^*$  is a relative price vector

$$\left( p_i^* \geq 0, \quad \sum_{i=1}^r p_i^* = 1 \right),$$

is a competitive equilibrium if:

$$U(\mathbf{x}^*) = \max_{\mathbf{x} \in \hat{X}} U(\mathbf{x}), \quad (9.1.3)$$

where  $\hat{X} = \{\mathbf{x} | \mathbf{x} \in X, (\mathbf{p}^*, \mathbf{x}) \leq (\mathbf{p}^*, \mathbf{y}^*)\}$ ;

$$(\mathbf{p}^*, \mathbf{y}^*) = \max_{\mathbf{y} \in Y} (\mathbf{p}^*, \mathbf{y}); \quad (9.1.4)$$

and

$$\mathbf{x}^* \leq \mathbf{y}^*, \quad (\mathbf{p}^*, \mathbf{x}^* - \mathbf{y}^*) = 0. \quad (9.1.5)$$

The key theorem relating competitive equilibrium and consumption of maximal utility is as follows.

► **THEOREM 9.1.1.** Let assumptions I and II be satisfied. A vector  $\mathbf{x}^*$  maximizes  $U(\mathbf{x})$  among all feasible vectors if and only if there exist a vector  $\mathbf{y}^*$  in  $Y$  and a price vector  $\mathbf{p}^*$  such that  $\{\mathbf{p}^*, \mathbf{x}^*, \mathbf{y}^*\}$  is a competitive equilibrium.

*Proof.* (The proof of this theorem is an application of Theorem 7.1.1.) Suppose that  $\mathbf{x}^*$  maximizes  $U(\mathbf{x})$  among all feasible vectors. Choose any  $\mathbf{y}^* \in Y$  such that  $\mathbf{y}^* \geq \mathbf{x}^*$ . Then  $\{\mathbf{x}^*, \mathbf{y}^*\}$  solves the problem of (9.1.1) and (9.1.2). Moreover, assumption I ensures that the extra hypothesis of Theorem 7.1.1 is satisfied. Consequently, by virtue of Theorem 7.1.1, the Lagrangian form

$$\phi(\mathbf{x}, \mathbf{y}; \mathbf{u}) = g(\mathbf{x}, \mathbf{y}) + (\mathbf{u}, \mathbf{F}(\mathbf{x}, \mathbf{y})) = U(\mathbf{x}) + (\mathbf{u}, \mathbf{F}(\mathbf{x}, \mathbf{y})),$$

where  $\{\mathbf{x}, \mathbf{y}\}$  traverses  $X \otimes Y$  and  $\mathbf{u} \geq \mathbf{0}$  has a saddle point  $\{\mathbf{x}^*, \mathbf{y}^*; \mathbf{u}^0\}$  such that

$$U(\mathbf{x}) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x}, \mathbf{y})) \leq U(\mathbf{x}^*) + (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^*, \mathbf{y}^*)) \leq U(\mathbf{x}^*) + (\mathbf{u}, \mathbf{F}(\mathbf{x}^*, \mathbf{y}^*)) \quad (9.1.6)$$

for all  $\mathbf{x}, \mathbf{y}$  in  $X \otimes Y$  and  $\mathbf{u} \geq \mathbf{0}$ . Furthermore, the same theorem indicates that

$$\mathbf{F}(\mathbf{x}^*, \mathbf{y}^*) \geq \mathbf{0}, \quad (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^*, \mathbf{y}^*)) = 0. \quad (9.1.7)$$

We now deduce that  $\mathbf{u}^0 > \mathbf{0}$ , since otherwise the left-hand inequality of (9.1.6) contradicts assumption II. Let  $\mathbf{p}^* = \lambda \mathbf{u}^0$ , where

$$\frac{1}{\lambda} = \sum_{i=1}^n u_i^0.$$

We now establish that  $\{\mathbf{x}^*, \mathbf{y}^*, \mathbf{p}^*\}$  constitutes a competitive equilibrium. First, note that (9.1.7) implies (9.1.5). Since

$$(\mathbf{u}^0, \mathbf{F}(\mathbf{x}, \mathbf{y})) = \frac{1}{\lambda} [(\mathbf{p}^*, \mathbf{y}) - (\mathbf{p}^*, \mathbf{x})],$$

we find with reference to (9.1.7) and the left-hand inequality of (9.1.6) that  $U(\mathbf{x}^*) \geq U(\mathbf{x})$  whenever  $\mathbf{x} \in \hat{X}$  in accordance with (9.1.3). Again, using the left-hand inequality of (9.1.6) for the special choice of  $\{\mathbf{x}, \mathbf{y}\} = \{\mathbf{x}^*, \mathbf{y}\}$ , with  $\mathbf{y}$  arbitrary in  $Y$ , we arrive at

$$(\mathbf{u}^0, \mathbf{y}) \leq (\mathbf{u}^0, \mathbf{y}^*),$$

which is equivalent to (9.1.4). The first half of the theorem has now been proved.

Suppose, conversely, that  $\{\mathbf{p}^*, \mathbf{x}^*, \mathbf{y}^*\}$  defines a competitive equilibrium in the sense of (9.1.3-5). We must prove that  $U(\mathbf{x}^*) = \max U(\mathbf{x})$  for all feasible  $\mathbf{x}$ . It is clear that for every feasible  $\mathbf{x}$  there exists some  $\mathbf{y}$  in  $Y$  such that  $\mathbf{y} \geq \mathbf{x}$ . Consequently,  $(\mathbf{p}^*, \mathbf{y}) \geq (\mathbf{p}^*, \mathbf{x})$ . Also  $(\mathbf{p}^*, \mathbf{y}^*) \geq (\mathbf{p}^*, \mathbf{x})$  by (9.1.4); and hence, by (9.1.3),  $U(\mathbf{x}^*) \geq U(\mathbf{x})$  as desired. This completes the proof.

*The case of many consumers.* Let  $X_i$  represent the consumption set in  $E'$  available to consumer  $i$  ( $i = 1, 2, \dots, m$ ), and let

$$X = \sum_{i=1}^m X_i$$

designate the cumulative consumption set of all consumers. The preference ordering of consumer  $i$  is described by a strictly concave utility function  $U_i(\mathbf{x}_i)$ . Each set  $X_i$  is assumed to be convex and closed, and lies in the nonnegative orthant of  $E'$ . The situation of the production possibility vectors is the same as in the case of a single consumer.

Again we seek to determine conditions which imply that the Pareto optimum is an element of a competitive equilibrium. A system of vectors  $\{\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_m^*; \mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_n^*; \mathbf{p}^*\}$  is said to be a competitive equilibrium if  $\mathbf{p}^*$  is a relative price vector

$$\left( p_i^* \geq 0, \quad \sum_{i=1}^m p_i^* = 1 \right)$$

such that

$$(\mathbf{p}^*, \mathbf{y}_j^*) = \max_{\mathbf{y}_j \in Y_j} (\mathbf{p}^*, \mathbf{y}_j); \quad (9.1.8)$$

$$U_i(\mathbf{x}_i^*) = \max_{\mathbf{x}_i \in X_i} U_i(\mathbf{x}_i), \quad (9.1.9)$$

where

$$\hat{X}_i = \left\{ \mathbf{x}_i \mid \mathbf{x}_i \in X_i, (\mathbf{p}^*, \mathbf{x}_i) \leq \sum_{j=1}^n \alpha_{ij} (\mathbf{p}^*, \mathbf{y}_j^*) \right\};^*$$

and

$$\mathbf{x}^* \leq \mathbf{y}^*, \quad (\mathbf{p}^*, \mathbf{x}^* - \mathbf{y}^*) = 0, \quad (9.1.10)$$

---

\* The term  $\alpha_{ij}$  is the share of the profit that is allotted to the  $i$ th consumer; it is assumed to satisfy the conditions

$$\sum_{i=1}^m \alpha_{ij} = 1 \quad (j = 1, \dots, m), \quad \alpha_{ij} \geq 0.$$

For the full economic interpretation of  $\alpha_{ij}$ , see Section 8.8.

where

$$\mathbf{x}^* = \sum_{i=1}^m \mathbf{x}_i^*, \quad \mathbf{y}^* = \sum_{j=1}^n \mathbf{y}_j^*.$$

We may formulate the problem of characterizing Pareto optima as a vector maximum problem in the sense of Theorem 7.4.1. Explicitly, we define for each  $\{\mathbf{x}_1, \dots, \mathbf{x}_m; \mathbf{y}_1, \dots, \mathbf{y}_n\}$  ( $\mathbf{x}_i \in X_i$ ;  $\mathbf{y}_j \in Y_j$ ) the vector function

$$\mathbf{G}(\mathbf{x}_1, \dots, \mathbf{x}_m; \mathbf{y}_1, \dots, \mathbf{y}_n) = \{U_i(\mathbf{x}_i)\} = \mathbf{U}(\mathbf{x}). \quad (9.1.11)$$

Let

$$\mathbf{F}(\mathbf{x}_1, \dots, \mathbf{x}_m; \mathbf{y}_1, \dots, \mathbf{y}_n) = \mathbf{y} - \mathbf{x}, \quad (9.1.12)$$

where

$$\mathbf{y} = \sum_{j=1}^n \mathbf{y}_j, \quad \mathbf{x} = \sum_{i=1}^m \mathbf{x}_i.$$

In accordance with the terminology of Section 7.4 the characterization of Pareto optima is equivalent to characterizing all efficient points of  $\{U_i(\mathbf{x}_i)\}$  subject to the constraints  $\mathbf{F} \geq \mathbf{0}$ ,  $\mathbf{x}_i \in X_i$ , and  $\mathbf{y}_j \in Y_j$ .

Again assumptions I and II hold, with the second extended in the natural manner to hold for each component  $U_i$ .

The relations between Pareto optimum vector systems and competitive equilibrium are derived from the statements of Theorem 7.4.1. The following theorem states these relations more precisely.

► **THEOREM 9.1.2.** Let assumptions I and II be fulfilled. Let

$$\{\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_m^*\}$$

be a Pareto optimum such that if  $\mathbf{y}^* \geq \mathbf{x}^*$  there exists a consumption vector system  $\{\tilde{\mathbf{x}}_i\}$  for each  $i_0$ , where  $\tilde{\mathbf{x}}_i = \mathbf{x}_i^*, i \neq i_0$ , and there is some  $\tilde{\mathbf{y}}$  in  $Y$  such that  $\tilde{\mathbf{y}} - \tilde{\mathbf{x}} \gg \mathbf{0}$ . Then there exist vectors  $\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_n^*$  ( $\mathbf{y}_i^* \in Y_i$ ), a price vector  $\mathbf{p}^*$ , and a matrix  $\{\alpha_{ij}^*\}$

$$\left( \alpha_{ij}^* \geq 0, \quad \sum_{i=1}^m \alpha_{ij}^* = 1 \quad \text{for all } j \right)$$

such that

$$\{\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_m^*; \mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_n^*; \mathbf{p}^*, \alpha_{ij}^*\}$$

is a competitive equilibrium.

*Proof.* Suppose that  $\{\mathbf{x}_1^*, \dots, \mathbf{x}_m^*\}$  is a Pareto optimum. Let  $\mathbf{y}_1^*, \dots, \mathbf{y}_n^*$  designate the respective terms of  $Y_i$ , giving

$$\mathbf{y}^* = \sum_{i=1}^n \mathbf{y}_i^*,$$

which is selected such that

$$\mathbf{y}^* \geq \mathbf{x}^* = \sum_{i=1}^m \mathbf{x}_i^*.$$

Clearly  $\{\mathbf{x}_1^*, \dots, \mathbf{x}_m^*\}$  is an efficient point of  $\mathbf{G}$  [equation (9.1.11)] among all vectors  $\mathbf{x}_i, \mathbf{y}_j$  satisfying the constraints  $\mathbf{F} = \mathbf{y} - \mathbf{x} \geq \mathbf{0}$ ,  $\mathbf{x}_i \in X_i$ , and  $\mathbf{y}_j \in Y_j$ . Because of assumption I the hypothesis of Theorem 7.4.1 is satisfied. By Theorem 7.4.1 there exist vectors  $\mathbf{u}^0$  and  $\mathbf{v}^0$  ( $\mathbf{u}^0 \geq \mathbf{0}; \mathbf{v}^0 > \mathbf{0}$ ) such that

$$\begin{aligned} (\mathbf{v}^0, \mathbf{U}(\mathbf{x})) + (\mathbf{u}^0, \mathbf{y} - \mathbf{x}) &\leq (\mathbf{v}^0, \mathbf{U}(\mathbf{x}^*)) + (\mathbf{u}^0, \mathbf{y}^* - \mathbf{x}^*) \\ &\leq (\mathbf{v}^0, \mathbf{U}(\mathbf{x}^*)) + (\mathbf{u}, \mathbf{y}^* - \mathbf{x}^*) \end{aligned} \quad (9.1.13)$$

for all  $\mathbf{u} \geq \mathbf{0}$  and  $\mathbf{x}_i \in X_i, \mathbf{y}_j \in Y_j$ . Since  $\mathbf{y}^* \geq \mathbf{x}^*$ , we conclude from the right-hand inequality that  $(\mathbf{u}^0, \mathbf{y}^* - \mathbf{x}^*) = 0$  (see the proof of Theorem 7.1.1). Next, observe that  $\mathbf{u}^0 > \mathbf{0}$ , since if  $\mathbf{u}^0 = \mathbf{0}$  and  $\mathbf{v}^0 > \mathbf{0}$  the left-hand inequality would be incompatible with assumption II. Finally, we observe that  $\mathbf{v}^0 \gg \mathbf{0}$ . For suppose to the contrary that  $v_{i_0}^0 = 0$ ; then in the left-hand inequality we may assign  $\mathbf{x}_i = \mathbf{x}_i^*$  for  $i \neq i_0$  (as guaranteed by the hypothesis) such that  $\mathbf{x} \ll \mathbf{y}$ . Inserting these choices of  $\mathbf{x}$  and  $\mathbf{y}$  into (9.1.13), we see plainly that the left-hand inequality is contradicted.

Define

$$\mathbf{p}^* = \frac{1}{\sum_{i=1}^m u_i^0} \mathbf{u}^0$$

and  $\alpha_{ij}^* = a_i$ , where

$$a_i = \frac{(\mathbf{p}^*, \mathbf{x}_i^*)}{(\mathbf{p}^*, \mathbf{y}^*)}.$$

Note that

$$a_i \geq 0, \quad \sum_{i=1}^m a_i = 1,$$

since

$$\sum_{i=1}^m (\mathbf{p}^*, \mathbf{x}_i^*) = (\mathbf{p}^*, \mathbf{y}^*), \quad (\mathbf{p}^*, \mathbf{x}_i) \geq 0.$$

We are now prepared to establish the claim that

$$\{\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_m^*; \quad \mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_n^*; \quad \mathbf{p}^*, \alpha_{ij}^*\}$$

is a competitive equilibrium.

Equation (9.1.10) follows immediately from

$$\mathbf{y}^* \geq \mathbf{x}^* \quad \text{and} \quad (\mathbf{u}^0, \mathbf{y}^* - \mathbf{x}^*) = 0.$$

Next, set  $\mathbf{x}_i = \mathbf{x}_i^*$  for all  $i$ , and  $\mathbf{y}_j = \mathbf{y}_j^*$  for  $j \neq j_0$ . Then the left-hand inequality of (9.1.13) yields

$$(\mathbf{u}^0, \mathbf{y}_{j_0}) \leq (\mathbf{u}^0, \mathbf{y}_{j_0}^*),$$

which gives (9.1.8) since  $j_0$  is arbitrary. Finally, set  $\mathbf{y}_j = \mathbf{y}_j^*$  for all  $j$  and  $\mathbf{x}_i = \mathbf{x}_i^*$  for  $i \neq i_0$ , and choose  $\mathbf{x}_{i_0}$  in  $\hat{X}_{i_0}$  [equation (9.1.9)]. The left-hand inequality of (9.1.13) then reduces to

$$v_{i_0}^0 U_{i_0}(\mathbf{x}_{i_0}) + \left( \sum_{i=1}^r u_i^0 \right) (\mathbf{p}^*, \mathbf{y}^* - \mathbf{x}) \leq v_{i_0}^0 U_{i_0}(\mathbf{x}_{i_0}^*). \quad (9.1.14)$$

Now  $\mathbf{x}_i^*$  belongs to  $\hat{X}_i$ , by the construction of  $\alpha_{ij}^*$ . Then,

$$\begin{aligned} (\mathbf{p}^*, \mathbf{x}_i^*) &= a_i(\mathbf{p}^*, \mathbf{y}^*) \quad (i \neq i_0), \\ (\mathbf{p}^*, \mathbf{x}_{i_0}) &\leq a_{i_0}(\mathbf{p}^*, \mathbf{y}^*), \end{aligned}$$

and by summation

$$(\mathbf{p}^*, \mathbf{x}) \leq (\mathbf{p}^*, \mathbf{y}^*).$$

From this observation, combined with  $v_{i_0}^0 > 0$ , we deduce that  $U_{i_0}(\mathbf{x}_{i_0}) \leq U_{i_0}(\mathbf{x}_{i_0}^*)$  for  $\mathbf{x}_{i_0}$  in  $\hat{X}_{i_0}$ , the statement of (9.1.9). This completes the proof of the theorem.

The restriction stated in the theorem was used in the preceding proof only in showing that  $v_i^0 > 0$  for all  $i$ . The reader will note that all conclusions of Theorem 9.1.2 follow whenever we can guarantee that  $\mathbf{v}^0 \gg \mathbf{0}$ . In this connection we refer to the discussion of Section 7.4, where alternative conditions are indicated which yield  $\mathbf{v}^0 \gg \mathbf{0}$ .

We turn to the converse of Theorem 9.1.2.

► **THEOREM 9.1.3.** If  $\{\mathbf{x}_1^*, \dots, \mathbf{x}_m^*; \mathbf{y}_1^*, \dots, \mathbf{y}_n^*; \mathbf{p}^*, \alpha_{ij}^*\}$  is a competitive equilibrium, then  $\{\mathbf{x}_1^*, \dots, \mathbf{x}_m^*\}$  is a Pareto optimum.

*Proof.* Suppose to the contrary that there exists a feasible vector system  $\{\mathbf{x}_i'\}$  such that  $U_i(\mathbf{x}_i') \geq U_i(\mathbf{x}_i^*)$  ( $i = 1, \dots, m$ ), with inequality for at least one  $i$ . Let us arrange the indices such that

$$\mathbf{x}_i' = \mathbf{x}_i^* \quad (i = 1, \dots, s), \quad \mathbf{x}_i' \neq \mathbf{x}_i^* \quad (i = s + 1, \dots, m).$$

Clearly  $s < m$ , since  $\mathbf{x}_i'$  dominates  $\mathbf{x}_i^*$  in the sense of utility by assumption. Since  $U_i$  is strictly concave, it follows that

$$\tilde{\mathbf{x}}_i = t\mathbf{x}_i' + (1 - t)\mathbf{x}_i^* \quad (i = 1, \dots, m; 0 < t < 1)$$

has

$$U_i(\tilde{\mathbf{x}}_i) > U_i(\mathbf{x}_i^*) \quad (i = s + 1, \dots, m).$$

Moreover,  $\tilde{\mathbf{x}}_i$  is feasible, and hence there exists  $\tilde{\mathbf{y}}_j$  such that

$$\tilde{\mathbf{y}} = \sum_{j=1}^n \tilde{\mathbf{y}}_j \geq \tilde{\mathbf{x}} = \sum_{i=1}^m \tilde{\mathbf{x}}_i.$$

Therefore  $(\mathbf{p}^*, \tilde{\mathbf{y}}) \geq (\mathbf{p}^*, \tilde{\mathbf{x}})$ , and by (9.1.8)

$$(\mathbf{p}^*, \mathbf{y}^*) \geq (\mathbf{p}^*, \tilde{\mathbf{x}}). \quad (9.1.15)$$

Taking into account  $\tilde{\mathbf{x}}_i = \mathbf{x}_i^*$  for  $i = 1, \dots, s$  and (9.1.15), we infer the existence of  $i_0$  ( $m \geq i_0 \geq s + 1$ ) such that

$$(\mathbf{p}^*, \tilde{\mathbf{x}}_{i_0}) \leq \sum_{j=1}^n \alpha_{i_0 j} (\mathbf{p}^*, \mathbf{y}_j^*).$$

However,  $U_{i_0}(\tilde{\mathbf{x}}_{i_0}) > U_{i_0}(\mathbf{x}_{i_0}^*)$ , in contradiction to (9.1.9). This completes the proof.

Now that we have essentially characterized the Pareto optimum consumption vector systems, the problem remains of how to single out a specific Pareto optimum. To do this, it is necessary to impose some further criterion. Many such criteria have been proposed. One of them, known as the principle of equity, amounts to assigning equal weights  $\lambda_i = 1/m$  and determining  $\mathbf{x}_i \in X_i$  such that

$$\sum_{i=1}^m \lambda_i U_i(\mathbf{x}_i)$$

is maximized, where the  $\mathbf{x}_i$  are subject to the constraints  $\mathbf{x}_i \in X_i$  and there exists some  $\mathbf{y}$  in  $Y$  for which

$$\mathbf{x} = \sum_{i=1}^m \mathbf{x}_i \leq \mathbf{y}.$$

The  $\mathbf{x}_i^*$  so obtained obviously constitutes a Pareto optimum. The economic merits of such a selection process have been argued cogently by Edgeworth and Marshall, among others.

In general, selecting a Pareto optimum represents a formidable philosophical problem very much akin to the task of justifiably specifying an admissible statistical procedure from a complete class of such procedures.

**9.2 The stability of a competitive equilibrium.** In the previous chapter and the preceding section, we were concerned primarily with investigating the existence, uniqueness, and optimality of a competitive equilibrium. These concepts are static. Of equal importance are the dynamics of equilibrium, and particularly the concept of stability. In this section we shall describe a formal dynamic model whose characteristics reflect the nature

of the competitive process, and examine its stability properties in the light of certain assumptions as to the properties of the individual units or of the aggregate excess demand functions.

To begin with, let us assume as before that there exist  $m$  consumers with available consumption sets  $X_i$  in  $E^{r+1}$  ( $i = 1, \dots, m$ ).† Consumer  $i$  seeks to choose a possible consumption vector  $\mathbf{x}_i$  from  $X_i$  which maximizes his utility (see p. 267). Let  $\mathbf{x} = \sum \mathbf{x}_i$  denote an aggregate demand consumption vector.

Let  $Y$  represent the aggregate production sector of commodity bundles in  $E^{r+1}$ . Each of the production units (firms) seeks to produce in a manner which maximizes profits. Consequently, with any price vector  $\mathbf{p} = \langle p_0, p_1, \dots, p_r \rangle$  of the commodity bundle there are associated an aggregate demand vector  $\mathbf{x}(\mathbf{p})$  and an aggregate production vector  $\mathbf{y}(\mathbf{p})$ , determined according to conditions (B) and (C) of p. 281, respectively, which possess the desired maximality characteristics. Subject to the usual conditions underlying equilibrium (primarily the dictum of unsaturated demand),

$$(\mathbf{p}, \mathbf{x}(\mathbf{p}) - \mathbf{y}(\mathbf{p})) = (\mathbf{p}, \mathbf{F}(\mathbf{p})) = 0 \quad (9.2.1)$$

for each price vector  $\mathbf{p}$ . For the derivation of (9.2.1) the reader should consult p. 284. We shall refer to  $\mathbf{F}(\mathbf{p})$  as the (aggregate) excess demand vector function for the price vector  $\mathbf{p}$ . The excess demand vector for the  $i$ th consumer is denoted by  $\mathbf{F}^i(\mathbf{p})$ ;

$$\mathbf{F}(\mathbf{p}) = \sum_{i=1}^m \mathbf{F}^i(\mathbf{p}).$$

It follows from the same considerations that  $(\mathbf{p}, \mathbf{F}^i(\mathbf{p})) = 0$ .

Throughout what follows we assume that  $\mathbf{F}(\mathbf{p})$  is continuous-differentiable and single-valued for  $\mathbf{p} > \mathbf{0}$ . Relation (9.2.1) is classically known as the Walras law. We emphasize that it operates universally, no matter what prices prevail in the market.

An equilibrium price vector  $\mathbf{p}^*$  is distinguished further by the property that

$$\mathbf{F}(\mathbf{p}^*) \leq \mathbf{0}, \quad (9.2.2)$$

which ensures that the consumption vector  $\mathbf{x}(\mathbf{p}^*)$  can be achieved by the production vector  $\mathbf{y}(\mathbf{p}^*)$ .

For mathematical convenience and to help clarify the dynamic price adjustment process, we shall assume in what follows that a strictly positive

---

† We are considering a commodity vector with  $r + 1$  components instead of  $r$  in order to be able to identify the zeroth component as the numéraire whenever it is important to distinguish it from the others.

equilibrium price vector  $\mathbf{p}^*$  exists, i.e., that  $\mathbf{p}^* \gg \mathbf{0}$ . This implies, in view of (9.2.1) and (9.2.2), that  $\mathbf{F}(\mathbf{p}^*) = \mathbf{0}$ .

We shall also assume that the excess demand function  $\mathbf{F}(\mathbf{p})$  is a positively homogeneous function of degree zero, i.e.,

$$\mathbf{F}(\lambda p_0, \lambda p_1, \dots, \lambda p_r) = \mathbf{F}(p_0, p_1, \dots, p_r) \quad (\lambda > 0). \quad (9.2.3)$$

This is the natural assumption that only the relative prices of the various commodities are relevant in influencing fluctuations of the excess demand functions.

We are now prepared to propose a dynamic model which describes how prices adjust over time to variations of supply and demand in a competitive economy.

One such dynamic model is based on the system of differential equations

$$\frac{dp_j}{dt} = F_j(p_0, p_1, \dots, p_r) \quad (j = 0, 1, \dots, r), \quad (9.2.4)$$

or in vector form

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}(\mathbf{p}). \quad (9.2.5)$$

For such a system the price of a commodity increases (decreases) when the excess demand for that commodity is positive (negative). The rate of increase or decrease is proportional to the amount of the excess demand or supply of each commodity. This appealing functional relationship between supply, demand, and prices is often postulated and represents the simplest and most direct characterization of the price adjustment process.

By a suitable choice of units of measurement for each of the commodities we can convert (9.2.4) into

$$\frac{dp_j}{dt} = k_j F_j \quad (j = 0, 1, \dots, r),$$

with the proportionality constants  $k_j$  positive and arbitrary. Thus there is no loss of generality in supposing  $k_j = 1$ .

A discrete-time analog of the dynamic system (9.2.4) is

$$p_j(t+1) = p_j(t) + \rho F_j[p_0(t), \dots, p_r(t)] \quad (j = 0, 1, \dots, r). \quad (9.2.6)$$

In view of the physical interpretation of the model, it is necessary for accuracy to indicate some restrictions in (9.2.5) and (9.2.6) which ensure that  $\mathbf{p}(t)$  is a nonnegative vector for all  $t$ .

In the case of discrete time we modify (9.2.6) to the system of equations

$$\mathbf{p}(t+1) = \max \{0, \mathbf{p}(t) + \rho \mathbf{F}[\mathbf{p}(t)]\}, \quad (9.2.7)$$

thus preserving the positivity of  $\mathbf{p}(t)$  for all  $t$  while still reflecting the dynamics of the price adjustment process. We postpone the analysis of (9.2.7) until Section 9.6.

For the case of continuous time we shall not bother to check whether  $\mathbf{p}(t)$  remains a positive vector at all times. The independent reader should be able to check this condition without too much difficulty.

We shall have occasion to study stability for two systems emerging from (9.2.5): (1) the normalized price adjustment process, (2) the nonnormalized price adjustment process. In the normalized process a single commodity (say the zeroth) is distinguished, and the price vector is normalized so that  $\mathbf{q} = \langle 1, q_1, \dots, q_r \rangle$ . The commodity singled out is frequently referred to as the numéraire. For the normalized process we consider the dynamics in terms of the relative prices  $q_j = p_j/p_0$  ( $j = 1, \dots, r$ ). Here we reformulate the price adjustment process and postulate that again it is described by a system of differential equations

$$\frac{dq_j}{dt} = G_j(\mathbf{q}) \quad (j = 1, \dots, r), \quad (9.2.8)$$

where  $G_j(\mathbf{q}) = F_j(1, q_1, \dots, q_r)$  and  $F_j$  has the same meaning as previously.

The nonnormalized adjustment process treats all commodities symmetrically, and the dynamics are expressed by means of (9.2.5).

Stability criteria will be established simultaneously for the normalized and nonnormalized processes. From a mathematical point of view the nonnormalized process is complicated by the multitude of equilibrium price vectors, since if  $\mathbf{p}^*$  is an equilibrium point, all points on the ray  $\lambda\mathbf{p}^*$  ( $\lambda > 0$ ) are equilibrium points. However, this model is the more appealing from an economic point of view because of its symmetry. The normalized process is more meaningful in many special situations, notably those in which the commodity "money" is not a free commodity and may therefore be distinguished as the numéraire.

A solution of (9.2.5) is denoted by  $\mathbf{p}(t) = \psi(t, \mathbf{p}^0)$ , where  $\mathbf{p}^0$  designates the initial price vector. The excess demand functions  $F_j(\mathbf{p})$  are henceforth assumed to possess enough smoothness properties to guarantee that the solution  $\psi(t, \mathbf{p}^0)$  is uniquely determined by, and changes continuously with, the initial price vector  $\mathbf{p}^0$ . A solution of (9.2.8) is denoted by  $\mathbf{q}(t) = \phi(t, \mathbf{q}^0)$ . Here we make similar assumptions regarding the uniqueness and continuity of solutions. Clearly, any strictly positive equilibrium vector

$$\mathbf{p}^*(t) = \psi(t, \mathbf{p}^*) = \mathbf{p}^*$$

is a fixed-point solution. With this background, the main problem of Sections 9.2 through 9.5 can be stated as follows: We seek to determine criteria

which imply that any solution  $\psi(t, \mathbf{p}^0)$  of (9.2.5) or solution  $\phi(t, \mathbf{p}^0)$  of (9.2.8) converges to an equilibrium vector.

Stability phenomena can be investigated on two levels. First, we can determine conditions for local stability, i.e., conditions implying convergence to the equilibrium vector  $\mathbf{p}^*$  when the initial price vector  $\mathbf{p}^0$  is sufficiently close to  $\mathbf{p}^*$ . Second, and more significantly, we can seek to determine conditions for global stability, i.e., conditions implying convergence to an equilibrium price vector regardless of the nature of the initial price vector  $\mathbf{p}^0$ .

At this point it seems worth while to summarize and discuss briefly the interpretation of various conditions which yield stability, reserving the detailed analysis for later sections.

Let

$$\mathbf{A}(\mathbf{p}) = \left\| \frac{\partial F_i}{\partial p_j} \right\|_{i,j=0, \dots, r} \quad \text{and} \quad \mathbf{B}(\mathbf{q}) = \left\| \frac{\partial G_i}{\partial q_j} \right\|_{i,j=1, \dots, r}. \quad (9.2.9)$$

We say that gross substitutability (strictly) prevails for the nonnormalized process if

$$\frac{\partial F_i}{\partial p_j} \geq 0 \quad (> 0) \quad \text{for every } i \neq j. \quad (9.2.10)$$

An analogous definition applies to the normalized process. Relation (9.2.10) has the property of gross substitutability for each  $\mathbf{p}$  if and only if  $\mathbf{A}(\mathbf{p})$  is an  $M$ -matrix (see p. 255). The economic justification underlying this property is as follows. If the price of the  $j$ th commodity rises while the prices of the other commodities remain unchanged, then an increase in the demand for every other commodity may be expected and hence (9.2.10) holds. Commodities related in this way are called "substitutes." This relationship is not always valid: if the price of butter goes up, for example, it may well happen that the demand for bread decreases, since many people cannot enjoy bread without butter.

Nevertheless, it is of some interest to study the price adjustment process in the situation of gross substitutability. Of particular interest is the principal theorem of Section 9.4, which states that if strict gross substitutability prevails for either the normalized or the nonnormalized process, and  $\mathbf{p}^*$  is a strictly positive equilibrium vector, then the process is globally stable.

Another proposition of some economic interest is that the normalized process (9.2.8) is globally stable whenever  $\mathbf{B}(\mathbf{q})$  defines a real negative quasi-definite matrix. (A real matrix  $\mathbf{C}$  is said to be negative quasi-definite if  $\mathbf{C} + \mathbf{C}'$  is negative definite.) The fact that  $\mathbf{B}(\mathbf{q})$  is a negative quasi-definite matrix is closely related to the concept of revealed preference [see Section 8.6 and equation (8.6.4)].

Other detailed results concerning stability are presented in the following sections.

**9.3 Local stability.** *Normalized process.* In studying local stability, we may replace the system (9.2.8) by a system of linear differential equations. Explicitly, by expanding the right-hand side of (9.2.8) about the equilibrium point  $\mathbf{q} = \mathbf{q}^*$  (throughout this section, unless there is an explicit statement to the contrary,  $\mathbf{q}^*$  is assumed to be  $\gg \mathbf{0}$ ) we obtain

$$\frac{d(\mathbf{q} - \mathbf{q}^*)}{dt} = \frac{d\mathbf{q}}{dt} \sim \mathbf{G}(\mathbf{q}^*) + \mathbf{B}(\mathbf{q}^*)(\mathbf{q} - \mathbf{q}^*) = \mathbf{B}(\mathbf{q}^*)(\mathbf{q} - \mathbf{q}^*), \quad (9.3.1)$$

where

$$\mathbf{B}(\mathbf{q}) = \left\| \frac{\partial G_i(\mathbf{q})}{\partial q_j} \right\|_{i,j=1, \dots, r}.$$

Local stability for the system (9.2.8) is obviously equivalent to the stability of the linear differential system

$$\frac{d\mathbf{z}}{dt} = \mathbf{C}\mathbf{z}, \quad (9.3.2)$$

where  $\mathbf{C} = \mathbf{B}(\mathbf{q}^*)$ , which in this circumstance reduces to the proposition that all solutions converge to zero. Clearly, the system (9.3.2) will be stable if and only if all characteristic roots of  $\mathbf{C}$  have negative real parts, since every solution of (9.3.2) is a linear combination of exponential functions  $e^{\lambda_i t}$ , where the  $\lambda_i$  represent eigenvalues of  $\mathbf{C}$ . A matrix  $\mathbf{C}$  whose eigenvalues all have negative real parts is called a stable matrix.

We now indicate a series of conditions which imply local stability. The hypotheses all possess some economic relevance.

► **THEOREM 9.3.1.** If  $\mathbf{B}(\mathbf{q}^*)$  is an  $M$ -matrix and either (a)  $\det \mathbf{B}(\mathbf{q}^*) \neq 0$  and  $\partial G_0 / \partial q_i \geq 0$  or (b)  $\partial G_0 / \partial q_i > 0$  ( $i = 1, 2, \dots, r$ ), then the normalized adjustment process (9.2.8) is locally stable.

(As already noted, the first assumption is synonymous with the statement that the normalized process locally (at  $\mathbf{q}^*$ ) has the property of gross substitutability.)

*Proof.* Differentiation of the Walras law [equation (9.2.1)] and evaluation at  $\mathbf{q}^*$  gives

$$\frac{\partial G_0(\mathbf{q}^*)}{\partial q_i} + \sum_{j=1}^r q_j^* \frac{\partial G_j(\mathbf{q}^*)}{\partial q_i} = 0. \quad (9.3.3)$$

Hence, by the hypothesis of our theorem,  $\mathbf{q}^* \mathbf{C} \leq \mathbf{0}$  and by assumption  $\mathbf{q}^* \gg \mathbf{0}$ . Since  $\mathbf{C}$  is an  $M$ -matrix, it is enough to show that  $\sigma(\mathbf{C}) < 0$ , where  $\sigma(\mathbf{C}) = \max \Re(\lambda_i)$  ( $\lambda_i$  are the eigenvalues of  $\mathbf{C}$ ). By adding a positive multiple of the identity we may suppose that  $\mathbf{D} = \mathbf{C} + \mu \mathbf{I}$  is a nonnegative matrix. Then

$$\mathbf{q}^* \mathbf{D} \leq \mu \mathbf{q}^*.$$

To complete the proof of the theorem it is necessary to verify that  $\lambda_0(\mathbf{D}) < \mu$ , where  $\lambda_0(\mathbf{D})$  is the spectral radius of  $\mathbf{D}$ .

Let  $\lambda_0$  denote the largest positive eigenvalue of  $\mathbf{D}$ , and suppose that  $\lambda_0 \geq \mu$ . In case (a),  $\lambda_0 > \mu$ . From the characterization of  $\lambda_0$  as given in Corollary 8.2.1, there exists a vector  $\mathbf{x} > \mathbf{0}$  such that  $\mathbf{D}\mathbf{x} > (\lambda_0 - \epsilon)\mathbf{x}$  for  $\lambda_0 - \epsilon > \mu$  and  $\epsilon$  small enough. But since  $\mathbf{q}^* \gg \mathbf{0}$ , it follows that

$$\mu(\mathbf{q}^*, \mathbf{x}) \geq (\mathbf{q}^*\mathbf{D}, \mathbf{x}) = (\mathbf{q}^*, \mathbf{D}\mathbf{x}) > (\lambda_0 - \epsilon)(\mathbf{q}^*, \mathbf{x}),$$

which is impossible.

In case (b), by (9.3.3)

$$\mathbf{q}^*\mathbf{D} \ll \mu\mathbf{q}^*.$$

Invoking the result of Corollary 8.2.2, we conclude that  $\lambda_0(\mathbf{D}) < \mu$  as desired. The proof of Theorem 9.3.1 is complete.

A further condition of the same kind which guarantees local stability is as follows.

►THEOREM 9.3.2. Let  $\mathbf{T} = \mathbf{B}(\mathbf{q}^*) + \mu\mathbf{I}$  be a positive matrix such that some power of  $\mathbf{T}$  is strictly positive. If  $\partial G_0/\partial q_i \geq 0$  ( $i = 1, 2, \dots, r$ ), with strict inequality for some  $i$ , then  $\mathbf{B}(\mathbf{q}^*)$  is stable.

(Note that the hypothesis is satisfied if  $\mathbf{B}(\mathbf{q}^*)$  has the property of strict gross substitutability.)

*Proof.* Relation (9.3.3) and the hypothesis together yield  $\mathbf{q}^*\mathbf{B}(\mathbf{q}^*) < \mathbf{0}$ , and hence  $\mathbf{q}^*\mathbf{T} < \mu\mathbf{q}^*$ . Some iterate of  $\mathbf{T}$  has  $\mathbf{q}^*\mathbf{T}^k \ll \mu^k\mathbf{q}^*$ . Consequently, by Corollary 8.2.2, the spectral radius of  $\mathbf{T}^k$  is smaller than  $\mu^k$ . A familiar result (see p. 383) tells us that the spectral radius of  $\mathbf{T}$  lies within the circle of radius  $\mu$ , and the theorem follows.

Other analytical criteria implying stability for matrices abound in the literature (see Problem 3 as typical).

We have concentrated here on developing theorems which indicate when an  $M$ -matrix is stable because of the natural appeal of the concept of “gross substitutes” and its transparent economic interpretations.

On the other end, two commodities labeled  $i$  and  $j$  are called complementary if  $\partial F_i/\partial p_j < 0$  and  $\partial F_j/\partial p_i < 0$ , e.g., pen and ink. If the price of ink rises, the demand for pens drops, and conversely.

Let us suppose that the indices of the matrix  $\mathbf{B}(\mathbf{q}^*)$  can be divided into two groups  $J$  and  $J'$  such that for two distinct indices  $k$  and  $k'$  in the same group (either  $J$  or  $J'$ )

$$b_{k,k'} \geq 0 \quad \text{and} \quad b_{k',k} \geq 0,$$

whereas for  $k$  in one group and  $k'$  in the other,

$$b_{k,k'} \leq 0 \quad \text{and} \quad b_{k',k} \leq 0.$$

We call such a matrix an  $\tilde{M}$ -matrix. For definiteness, let us arrange the rows and columns so that the indices of  $J$  precede the indices of  $J'$ . Clearly, if  $\mathbf{C}$  is an  $\tilde{M}$ -matrix, then  $\mathbf{SCS}$  is an  $M$ -matrix, where

$$\mathbf{S} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix},$$

with the identity matrix  $\mathbf{I}$  of the same order as the number of indices in  $J$  and  $-\mathbf{I}$  of the same order as the number of indices in  $J'$ .

► **THEOREM 9.3.3.** If  $\mathbf{B}(\mathbf{q}^*)$  is an  $\tilde{M}$ -matrix and there exists a vector  $\mathbf{u}$  such that  $\mathbf{v} = \mathbf{Su} > \mathbf{0}$  and  $\mathbf{SB}(\mathbf{q}^*)\mathbf{u} \ll \mathbf{0}$ , then  $\mathbf{B}(\mathbf{q}^*)$  is stable.

*Proof.* Since  $\mathbf{SS} = \mathbf{I}$ , we have

$$\mathbf{0} \gg \mathbf{SB}(\mathbf{q}^*)\mathbf{u} = \mathbf{SB}(\mathbf{q}^*)\mathbf{SSu} = \mathbf{Dv},$$

where  $\mathbf{D}$  is an  $M$ -matrix and  $\mathbf{v} > \mathbf{0}$ . It follows from Problem 15 of Chapter 8 that  $\mathbf{D}$  is stable; and therefore  $\mathbf{B}(\mathbf{q}^*)$ , being obtained from  $\mathbf{D}$  by a similarity transformation, is also stable. (See also Problem 4.)

*Nonnormalized process.* Stability for the nonnormalized process must be interpreted, on account of the homogeneity of the excess demand functions  $F_i(\mathbf{p})$ ; as noted earlier, the whole ray  $\Gamma = \{\lambda\mathbf{p}^*\} (\lambda > 0)$  consists of equilibrium vectors. In the following discussion, local stability shall mean that if  $\mathbf{p}$  is a price vector near the ray  $\Gamma$ , then  $\mathbf{p}(t)$  tends to some equilibrium vector situated on the ray  $\Gamma$ .

► **THEOREM 9.3.4.** Let  $\mathbf{A}(\mathbf{p}^*)$  be a strict  $M$ -matrix (that is,  $a_{ij} > 0; i \neq j$ ), with  $i, j = 0, 1, \dots, r$ . If  $\mathbf{p}^* \gg \mathbf{0}$  is an equilibrium vector and  $\mathbf{p}$  is sufficiently close to  $\Gamma = \{\lambda\mathbf{p}^* | \lambda > 0\}$ , then  $\mathbf{p}(t)$  converges to an equilibrium point of  $\Gamma$ .

The proof depends on the following lemma.

**LEMMA 9.3.1.** Given the hypothesis of Theorem 9.3.4, if  $\mathbf{p}$  is sufficiently close to  $\Gamma$ , then

$$(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) > \mathbf{0} \quad \text{for} \quad \mathbf{p} \neq \lambda\mathbf{p}^*.$$

*Proof.* Let  $h(\theta) = (\mathbf{p}^*, \mathbf{F}[\mathbf{p}(\theta)])$ , where  $\mathbf{p}(\theta) = \mathbf{p}^* + \theta(\mathbf{p} - \mathbf{p}^*)$ ; then we shall prove that  $h(0) = h'(0) = 0$  and  $h''(0) > 0$  unless  $\mathbf{p} = \lambda\mathbf{p}^*$  for some positive  $\lambda$ .

The conclusion  $h(0) = 0$  is immediate from the Walras law (9.2.1). By differentiation of (9.2.1), since  $\mathbf{p}^*$  is an equilibrium vector [ $\mathbf{F}(\mathbf{p}^*) = \mathbf{0}$ ], we deduce that  $h'(0) = 0$ . Explicitly,

$$h'(\theta) = \sum_{i=0}^r p_i^* \sum_{j=0}^r \frac{\partial F_i[\mathbf{p}(\theta)]}{\partial p_j} (p_j - p_j^*). \quad (9.3.4)$$

By differentiation of (9.2.1) with respect to  $p_j$  we obtain

$$\sum_{i=0}^r p_i \frac{\partial F_i(\mathbf{p})}{\partial p_j} + F_j(\mathbf{p}) = 0. \quad (9.3.5)$$

At  $\mathbf{p} = \mathbf{p}^*$  this reduces to  $\mathbf{p}^* \mathbf{A}(\mathbf{p}^*) = \mathbf{0}$ . Using this in (9.3.4), after changing the order of summation, shows that  $h'(0) = 0$ .

By differentiation of (9.3.4) and the same manipulations with (9.3.5), we obtain

$$h''(0) = - \sum_{j,k=0}^r (p_j - p_j^*) \left[ \frac{\partial F_j(\mathbf{p}^*)}{\partial p_k} + \frac{\partial F_k(\mathbf{p}^*)}{\partial p_j} \right] (p_k - p_k^*), \quad (9.3.6)$$

or in vector inner product notation

$$h''(0) = -(\mathbf{p} - \mathbf{p}^*, [\mathbf{A}(\mathbf{p}^*) + \mathbf{A}'(\mathbf{p}^*)](\mathbf{p} - \mathbf{p}^*)) = -(\mathbf{p} - \mathbf{p}^*, \mathbf{T}(\mathbf{p} - \mathbf{p}^*)),$$

where  $\mathbf{A}'$  as usual denotes the transpose of  $\mathbf{A}$ . We have already noted that  $\mathbf{p}^* \mathbf{A}(\mathbf{p}^*) = \mathbf{0}$ . The homogeneity of  $F_i(\mathbf{p})$ , which is equivalent to the Euler relation

$$\sum_{j=0}^r \frac{\partial F_i(\mathbf{p})}{\partial p_j} p_j = 0,$$

gives in particular  $\mathbf{A}(\mathbf{p}^*) \mathbf{p}^* = \mathbf{0}$ . It follows that  $\mathbf{p}^* [\mathbf{A}(\mathbf{p}^*) + \mathbf{A}'(\mathbf{p}^*)] = \mathbf{p}^* \mathbf{T} = \mathbf{0}$ . But  $\mathbf{T}$  is a real symmetric  $M$ -matrix, and since  $\mathbf{p}^* \gg \mathbf{0}$  we may invoke property (b) of  $M$ -matrices to conclude that any other solution  $\mathbf{x}$  of  $\mathbf{x} \mathbf{T} = \mathbf{0}$  is a scalar multiple of  $\mathbf{p}^*$ , and that all nonzero eigenvalues of  $\mathbf{T}$  are negative. Thus,  $\mathbf{T}$  is negative semidefinite. Hence  $h''(0) > 0$  unless  $\mathbf{p} = \lambda \mathbf{p}^*$ , again by virtue of the fact that  $F_i$  is homogeneous. The result of Lemma 9.3.1 is now clear, provided that we alter  $\mathbf{p}^*$  to a suitable  $\lambda \mathbf{p}^*$  ( $\lambda > 0$ ) in  $(\mathbf{p}^*, \mathbf{F}(\mathbf{p}))$  and form a Taylor expansion about  $\lambda \mathbf{p}^*$ .

*Proof of Theorem 9.3.4.* Define

$$V(t) = \frac{1}{2} |\mathbf{p}(t) - \mathbf{p}^*|^2 = \frac{1}{2} \sum_{i=0}^r [p_i(t) - p_i^*]^2,$$

where  $\mathbf{p}(t) = \psi(t, \mathbf{p}^0)$  and  $\mathbf{p}^0$  is sufficiently close to the ray  $\lambda \mathbf{p}^*$ . We calculate

$$\frac{dV(t)}{dt} = (\mathbf{p}(t) - \mathbf{p}^*, \mathbf{F}[\mathbf{p}(t)]) = -(\mathbf{p}^*, \mathbf{F}[\mathbf{p}(t)]),$$

the second equality resulting from the Walras law. By Lemma 9.3.1,  $dV(t)/dt < 0$  unless  $\mathbf{p}(t) = \lambda \mathbf{p}^*$ . Since  $V(t)$  decreases, we may select a limit point of  $\mathbf{p}(t) = \tilde{\mathbf{p}}$ . If  $\tilde{\mathbf{p}}$  is not on  $\Gamma$ , an easy contradiction ensures, as

follows. There exists a sequence  $t^\nu \rightarrow \infty$  such that

$$\lim_{\nu \rightarrow \infty} \psi(t^\nu, \mathbf{p}^0) = \tilde{\mathbf{p}}.$$

The continuity of solutions with respect to the initial position vector clearly implies that

$$\lim_{\nu \rightarrow \infty} \psi(t^\nu + t, \mathbf{p}^0) = \psi(t, \tilde{\mathbf{p}}).$$

Now  $\frac{dV(t^\nu + t)}{dt}$  tends to zero, but if  $\tilde{\mathbf{p}}$  is not on  $\Gamma$ ,

$$\lim_{\nu \rightarrow \infty} \frac{dV(t^\nu + t)}{dt} = \frac{d}{dt} (\frac{1}{2} |\psi(t, \tilde{\mathbf{p}}) - \mathbf{p}^*|^2) < 0.$$

Consequently, if  $\lim V(t) = a$ , then all limit points of  $\mathbf{p}(t)$  lie on the ray  $\Gamma$  at distance  $2a$  from  $\mathbf{p}^*$ . Any  $\tilde{\mathbf{p}}$  consists of at most one of two points. However, the differential equation (9.2.5) shows, with the aid of (9.2.1) that

$$\sum_{i=0}^r p_i(t) \frac{dp_i}{dt} = \sum_{i=0}^r p_i(t) F_i[\mathbf{p}(t)] = 0,$$

and hence

$$\sum_{i=0}^r p_i^2(t)$$

is a constant  $b$  independent of  $t$ . Clearly,  $\tilde{\mathbf{p}} = \lambda \mathbf{p}^*$ , where

$$\lambda = \sqrt{\frac{b}{\sum_{i=0}^r p_i^{*2}}}.$$

**\*9.4 Global stability of price adjustment processes.** In this section we investigate various conditions which imply stability in the large. The criteria are mostly based on economic considerations. We concentrate on the nonnormalized process whose dynamics are described by the differential-equations system (9.2.5).

In this section again (unless there is an explicit statement to the contrary) we assume the validity of the Walras law, the homogeneity of the excess demand functions  $F_i$ , and the strict positivity of an equilibrium price vector  $\mathbf{p}^*$ , which is postulated to exist (see also Problem 13). The following lemma is crucial to most of our discussion of global stability.

► **LEMMA 9.4.1.** If  $(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) > 0$  for all vectors  $\mathbf{p}$  for which  $\mathbf{F}(\mathbf{p}) \neq \mathbf{0}$ , then the nonnormalized price adjustment process is globally stable; i.e., as  $t$  tends to infinity the solution  $\psi(t, \mathbf{p}^0)$  converges to a fixed-point solution.

After proving Lemma 9.4.1, we shall devote the remainder of this section to establishing natural conditions which imply the hypothesis of this lemma. Theorems 9.4.1 and 9.4.2 state two such conditions.

The literature on the global stability of differential-equations systems is based largely on one fundamental method of analysis. This method consists in introducing some appropriate norm of the distance between the solution and the fixed point. If it can be shown that the values of this norm strictly decrease with time whenever the distance is not zero, then stability in the large follows. The norm we use in this section is the ordinary Euclidean distance. In the following section we employ two other norms that are suitable for the hypothesis proposed there (see also the notes at the close of the chapter). A more general technique for establishing stability is to consider functions  $V[\mathbf{p}(t)]$  which are not necessarily norms but possess appropriate contraction properties relative to equilibrium points. We exploit this concept on pp. 316–325.

*Proof of Lemma 9.4.1.* Let  $\mathbf{p}^0$  be any initial vector and let  $\mathbf{p}(t)$  designate the solution of (9.2.5) with initial position  $\mathbf{p}^0$ ; that is,  $\mathbf{p}(t) = \psi(t, \mathbf{p}^0)$ . Consider the norm deviation measure

$$V(t) = \frac{1}{2}|\mathbf{p}(t) - \mathbf{p}^*|^2 = \frac{1}{2}(\mathbf{p}(t) - \mathbf{p}^*, \mathbf{p}(t) - \mathbf{p}^*).$$

It follows from (9.2.5) that

$$V'(t) = (\mathbf{p}(t) - \mathbf{p}^*, \mathbf{F}[\mathbf{p}(t)]) = -(\mathbf{p}^*, \mathbf{F}[\mathbf{p}(t)]), \quad (9.4.1)$$

the last simplification resulting from an application of the Walras law. By the hypothesis,  $V'(t) < 0$  unless  $\mathbf{p}(t)$  satisfies  $\mathbf{F}(\mathbf{p}(t)) = \mathbf{0}$ . The proof from here proceeds as in the argument of Theorem 9.3.4.

We now discuss a series of applications of Lemma 9.4.1.

The aggregate excess demand functions are said to satisfy the weak axiom of revealed preference if

$$(\mathbf{p}', \mathbf{F}(\mathbf{p}'') - \mathbf{F}(\mathbf{p}')) \leq 0 \quad \text{and} \quad \mathbf{F}(\mathbf{p}'') \not\leq \mathbf{F}(\mathbf{p}') \quad (9.4.2)$$

imply

$$(\mathbf{p}'', \mathbf{F}(\mathbf{p}'') - \mathbf{F}(\mathbf{p}')) < 0 \quad (\text{see also p. 269}).$$

► **THEOREM 9.4.1.** If the excess demand functions satisfy the weak axiom of revealed preference, then the process described by (9.2.5) is globally stable.

*Proof.* Let  $\mathbf{p}'' = \mathbf{p}^*$  and  $\mathbf{p}' = \mathbf{p}$ . Then the first inequality of (9.4.2) is fulfilled because of the Walras law and by the definition of an equilibrium point  $[\mathbf{F}(\mathbf{p}^*) = \mathbf{0}]$ . Also, if  $\mathbf{p}$  is not an equilibrium price vector, we have

$\mathbf{F}(\mathbf{p}) \neq \mathbf{0}$  and hence the second inequality of (9.4.2) is also satisfied. It follows that  $-(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) < 0$ , which means that the hypothesis of Lemma 9.4.1 holds. Consequently, the process is globally stable and the theorem is established.

The weakness of Theorem 9.4.1 lies in the fact that the axiom of revealed preference is usually verified only in the neighborhood of an equilibrium point and need not hold in the large. The following result is deeper and also has considerable economic appeal.

► **THEOREM 9.4.2.** If the excess demand functions satisfy the property of strict gross substitutability in the interior of the positive orthant, then the process described by (9.2.5) is globally stable. Moreover,  $(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) > \mathbf{0}$  for all  $\mathbf{p} \neq \lambda \mathbf{p}^*$  ( $\lambda$  real and positive), where  $\mathbf{p}^* \gg \mathbf{0}$  is an equilibrium vector.

*Remarks on the hypothesis of Theorem 9.4.2.*† It must be emphasized that the restriction of strict gross substitutability to the interior of the positive orthant, as opposed to the full closed positive orthant, is necessary. Indeed, the property of strict gross substitutability and the requirement of homogeneity are contradictory. For let us assume to the contrary that the excess demand function  $\mathbf{F}(\mathbf{p})$  is defined and homogeneous and satisfies the property of strict gross substitutes for all

$$\mathbf{p} = \langle p_0, p_1, \dots, p_r \rangle \quad (p_i \geq 0),$$

including the boundary. Now, let  $\mathbf{p}^0 = \langle 0, p_1^0, \dots, p_r^0 \rangle$ , with  $\mathbf{p}^0 > \mathbf{0}$ . Then  $F_0(\lambda \mathbf{p}^0) = F_0(\mathbf{p}^0)$  is a constant in  $\lambda$ . But at least one of the components  $\lambda p_1^0, \dots, \lambda p_r^0$  strictly increases, and hence  $F_0$  should strictly increase, a contradiction.

If strict gross substitutability is assumed merely for the interior of the positive orthant, then the assumptions of homogeneity, the validity of the Walras law, and the property of strict gross substitutability are fully consistent.

Examples of vector functions  $\mathbf{F}$  obeying the stated restrictions are constructed as follows. For  $\mathbf{p} \gg \mathbf{0}$ , we define

$$F_k(\mathbf{p}) = \frac{\sum_{i=0}^r a_{ik} p_i}{p_k} \quad (k = 0, \dots, r). \quad (9.4.3)$$

The further conditions  $a_{ik} > 0$  for  $i \neq k$  and

$$\sum_{k=0}^r a_{ik} = 0$$

---

† It is indicated in Problem 7 that subject to the conditions of the theorem, any solution of (9.2.5) with  $\mathbf{p}(0) \gg \mathbf{0}$  necessarily has  $\mathbf{p}(t) \gg \mathbf{0}$  for all  $t > 0$ , and thus  $\mathbf{F}[\mathbf{p}(t)]$  is well defined.

imply trivially that the property of strict gross substitutability prevails and that the Walras law holds. That  $F_k(\mathbf{p})$  are homogeneous functions is also immediately clear. A simple special instance of (9.4.3) is obtained by setting  $a_{ik} = 1$  for  $i \neq k$  and  $a_{kk} = -r$ , ( $i, k = 0, 1, \dots, n$ ).

It is interesting to note that an excess demand function of the structure (9.4.3) can be derived by performing an appropriate consumer utility maximization. Explicitly, the problem

$$U(\mathbf{x}^*) = \max_{\mathbf{x}} U(\mathbf{x}) = \max_{\mathbf{x}_i} \sum_{i=0}^r \alpha_i \log (x_i + x_i^0 + c_i)$$

subject to the constraint

$$\sum_{i=0}^r p_i x_i = 0,$$

which is easily resolved by the standard Lagrange multiplier methods, leads to  $x_i^*(\mathbf{p}) = F_i(\mathbf{p})$  of the form (9.4.3). (The  $\alpha_i$  are fixed positive constants and  $\mathbf{x}^0 \gg \mathbf{0}$  is interpreted as a fixed vector of initial holdings for an individual consumer.)

If to the hypotheses of the theorem we add the condition that  $F_j(\mathbf{p})$  tends to infinity whenever  $\mathbf{p}$  approaches a vector  $\mathbf{p}^0 > \mathbf{0}$  for which  $p_j^0 = 0$  and otherwise approaches a finite limit, we may secure the existence of an equilibrium price vector  $\mathbf{p}^* \gg \mathbf{0}$  (see Problems 12 and 13 and the discussion of solutions to the problems on p. 357).

*Proof of Theorem 9.4.2.* We shall prove that if  $\mathbf{p} \neq \lambda \mathbf{p}^*$ , then

$$(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) > 0, \quad (9.4.4)$$

from which, in view of Lemma 9.4.1, the theorem follows.

We may without loss of generality assume that

$$\mathbf{p}^* = \mathbf{u}^* = \langle 1, 1, \dots, 1 \rangle.$$

To demonstrate this, we let

$$H_i(\mathbf{p}) = H_i(p_0, p_1, \dots, p_r) = p_i^* F_i(p_0^* p_0, p_1^* p_1, p_2^* p_2, \dots, p_r^* p_r).$$

It is readily verified that the vector function  $\mathbf{H}$  satisfies all the properties that  $\mathbf{F}$  satisfies, including gross substitutability; but now  $\mathbf{H}(\mathbf{u}^*) = \mathbf{0}$  by the Walras law, and thus  $\mathbf{u}^*$  is an equilibrium vector for  $\mathbf{H}$ . Finally, we observe that  $(\mathbf{u}^*, \mathbf{H}(\mathbf{p})) > 0$  for  $\mathbf{p} \neq \lambda \mathbf{u}^*$  if and only if  $(\mathbf{p}^*, \mathbf{F}(\mathbf{p})) > 0$  for  $\mathbf{p} \neq \lambda \mathbf{p}^*$ .

At this point it is convenient to revert to our old notation involving the vector function  $\mathbf{F}$  and the equilibrium vector  $\mathbf{p}^* = \langle 1, 1, \dots, 1 \rangle$ .

Let  $\mathbf{p} = \langle p_0, p_1, \dots, p_r \rangle$  denote any vector not a multiple of  $\mathbf{p}^*$ . Without loss of generality we may so number the commodities that

$$p_0 \leq p_1 \leq \dots \leq p_r. \quad (9.4.5)$$

Since  $\mathbf{p}$  is not a multiple of  $\mathbf{p}^*$ , there must be at least one case of strict inequality in (9.4.5), say  $p_\nu < p_{\nu+1}$  for some  $0 \leq \nu < r$ . We now define a sequence of vectors  $\mathbf{p}^0, \mathbf{p}^1, \dots, \mathbf{p}^r$  by the relations

$$\begin{aligned} \mathbf{p}^0 &= \langle p_0, p_0, p_0, \dots, p_0 \rangle, \\ \mathbf{p}^1 &= \langle p_0, p_1, p_1, \dots, p_1 \rangle, \\ &\dots \\ \mathbf{p}^s &= \langle p_0, p_1, \dots, p_{s-1}, p_s, p_s, \dots, p_s \rangle, \\ &\dots \\ \mathbf{p}^r &= \langle p_0, p_1, \dots, p_r \rangle, \end{aligned} \quad (9.4.6)$$

where the  $p_j$  are the components of  $\mathbf{p}$ . Obviously  $\mathbf{p}^r = \mathbf{p}$  and  $\mathbf{p}^0$  is a multiple of  $\mathbf{p}^*$ . We shall prove that

$$(\mathbf{p}^*, \mathbf{F}(\mathbf{p}^{s+1}) - \mathbf{F}(\mathbf{p}^s)) \geq 0 \quad (s = 0, 1, \dots, r-1), \quad (9.4.7)$$

with strict inequality for  $s = \nu$ . Adding the equations (9.4.7) and recognizing that  $\mathbf{F}(\mathbf{p}^0) = \mathbf{0}$ , we deduce the desired inequality (9.4.4).

It remains to prove the correctness of (9.4.7). Because of the assumption of strict gross substitutability, since only the components greater than  $s$  increase, it follows that

$$F_i(\mathbf{p}^{s+1}) - F_i(\mathbf{p}^s) \geq 0 \quad (i = 0, 1, \dots, s), \quad (9.4.8)$$

where strict equality holds when  $s = \nu$ .

Moreover, the homogeneity of  $\mathbf{F}$ , in conjunction with the fact that  $p_s/p_{s+1} \leq 1$ , shows that

$$\begin{aligned} F_j(\mathbf{p}^{s+1}) - F_j(\mathbf{p}^s) &= F_j \left( \frac{p_s}{p_{s+1}} p_0, \frac{p_s}{p_{s+1}} p_1, \dots, \frac{p_s}{p_{s+1}} p_s, p_s, p_s, \dots, p_s \right) \\ &\quad - F_j(p_0, p_1, \dots, p_{s-1}, p_s, p_s, \dots, p_s) \leq 0 \end{aligned} \quad (9.4.9)$$

for  $j = s+1, \dots, r$ . This follows from the hypothesis of strict gross substitutability, on the assumption that  $F_j$  is first evaluated at a price vector whose initial  $s+1$  components do not exceed in magnitude the respective values of the first  $s+1$  components of the second vector at which  $F_j$  is evaluated, while the last  $r-(s+1)$  coordinates are identical. Again strict inequality prevails in (9.4.9) if  $s = \nu$ .

For fixed  $j$ , iterating (9.4.9) to the zero index shows that

$$F_j(\mathbf{p}^{s+1}) \leq F_j(\mathbf{p}^s) \leq \cdots \leq F_j(\mathbf{p}^0) = 0 \quad (j = s+1, \dots, r), \quad (9.4.10)$$

with strict inequality at the point  $\mathbf{p}^r$  in the sequence (9.4.10), provided that  $\mathbf{p}^r$  appears.

Consider any fixed index  $s$  and partition all vectors of  $r+1$  components into two smaller vectors, the first consisting of the first  $s+1$  ordered components and the second consisting of the remaining components. If  $\mathbf{x}$  is a general vector of  $r+1$  components, we write  $\mathbf{x}_*$  and  $\mathbf{x}_{**}$  for the two associated vectors constructed by the rules just set forth. Note that

$$\mathbf{p}_*^{s+1} = \mathbf{p}_*^s \quad (9.4.11)$$

and

$$\mathbf{p}_{**}^{s+1} = (p_r^{s+1})\mathbf{p}_{**}^* = \mathbf{p}_{**}^s + (p_{s+1} - p_s)\mathbf{p}_{**}^*, \quad (9.4.12)$$

where  $\mathbf{p}_{**}^*$  is a vector of appropriate size with all components 1. Decomposing the expression  $(\mathbf{p}^{s+1}, \mathbf{F}(\mathbf{p}^{s+1})) - (\mathbf{p}^s, \mathbf{F}(\mathbf{p}^s))$  (which is zero by the Walras law) in accordance with the partitioning scheme suggested above, we obtain

$$\begin{aligned} 0 &= (\mathbf{p}_*^{s+1}, \mathbf{F}(\mathbf{p}^{s+1})_*) + (\mathbf{p}_{**}^{s+1}, \mathbf{F}(\mathbf{p}^{s+1})_{**}) \\ &\quad - (\mathbf{p}_*^s, \mathbf{F}(\mathbf{p}^s)_*) - (\mathbf{p}_{**}^s, \mathbf{F}(\mathbf{p}^s)_{**}) \\ &= (\mathbf{p}_*^s, \mathbf{F}(\mathbf{p}^{s+1})_* - \mathbf{F}(\mathbf{p}^s)_*) + (\mathbf{p}_{**}^{s+1}, \mathbf{F}(\mathbf{p}^{s+1})_{**}) - (\mathbf{p}_{**}^s, \mathbf{F}(\mathbf{p}^s)_{**}) \\ &\quad \quad \quad [\text{by (9.4.11)}] \\ &= (\mathbf{p}_*^s, \mathbf{F}(\mathbf{p}^{s+1})_* - \mathbf{F}(\mathbf{p}^s)_*) + (\mathbf{p}_{**}^{s+1}, \mathbf{F}(\mathbf{p}^{s+1})_{**} - \mathbf{F}(\mathbf{p}^s)_{**}) \\ &\quad \quad \quad + (p_{s+1} - p_s)(\mathbf{p}_{**}^*, \mathbf{F}(\mathbf{p}^s)_{**}) \\ &\quad \quad \quad [\text{by (9.4.12) and rearrangement}] \\ &= (\mathbf{p}_*^s, \mathbf{F}(\mathbf{p}^{s+1})_* - \mathbf{F}(\mathbf{p}^s)_*) + p_r^{s+1}(\mathbf{p}_{**}^*, \mathbf{F}(\mathbf{p}^{s+1})_{**} - \mathbf{F}(\mathbf{p}^s)_{**}) \\ &\quad \quad \quad + (p_{s+1} - p_s)(\mathbf{p}_{**}^*, \mathbf{F}(\mathbf{p}^s)_{**}) \quad [\text{by (9.4.12)}] \\ &\leq p_r^{s+1}(\mathbf{p}^*, \mathbf{F}(\mathbf{p}^{s+1}) - \mathbf{F}(\mathbf{p}^s)) + (p_{s+1} - p_s)(\mathbf{p}_{**}^*, \mathbf{F}(\mathbf{p}^s)_{**}) \\ &\quad \quad \quad [\text{by (9.4.5) and (9.4.11)}]. \end{aligned}$$

This may be written as

$$(\mathbf{p}^*, \mathbf{F}(\mathbf{p}^{s+1}) - \mathbf{F}(\mathbf{p}^s)) \geq -\frac{p_{s+1} - p_s}{p_r^{s+1}} (\mathbf{p}_{**}^*, \mathbf{F}(\mathbf{p}^s)_{**}). \quad (9.4.13)$$

The right-hand side of (9.4.13) is nonnegative because

$$\frac{p_{s+1} - p_s}{p_r^{s+1}} \geq 0$$

and  $(\mathbf{p}^*, \mathbf{F}(\mathbf{p}^*)) \leq 0$ . Furthermore, strict inequality holds for  $s = v$ , in view of (9.4.10) and the fact that  $p_{v+1} > p_v$ . Hence (9.4.7) follows and the proof is complete.

The global stability of the system (9.2.5) under the conditions of Theorem 9.4.2 can be established more simply as indicated in Theorem 9.4.3 below. However, the interesting supplementary inequality (9.4.4) does not appear to be susceptible to easy verification except by the reasoning of the above proof.

We next consider a price adjustment process defined by the system of differential equations

$$\frac{dp_i}{dt} = H_i[F_i(\mathbf{p})], \quad (i = 0, 1, \dots, r), \quad (9.4.14)$$

where  $H_i(x)$  is continuous and sign-preserving, i.e.,  $\text{sign } H_i(x) = \text{sign } x$  for all real  $x$ . The usual assumptions regarding uniqueness and continuity of the solution of (9.4.14) as a function of the initial position are made.

► **THEOREM 9.4.3.** Under the conditions of Theorem 9.4.2, with the single change that the Walras law is not required to hold, any solution  $\mathbf{p}(t)$  satisfying (9.4.14) with  $\mathbf{p}(0) \gg \mathbf{0}$  is stable.

*Proof.* As demonstrated in the proof of Theorem 9.4.2, we may without loss of generality normalize any strict positive equilibrium vector to the form  $\mathbf{p}^* = \langle 1, \dots, 1 \rangle$ . We suppose henceforth that this is done.

The following auxiliary result is needed.

(I) If  $\mathbf{p} = \langle p_0, \dots, p_r \rangle$  is strictly positive and satisfies

$$p_{i_0} = \max_{0 \leq j \leq r} p_j > \min_{0 \leq j \leq r} p_j = p_{k_0}$$

then  $F_{i_0}(\mathbf{p}) < 0$  and  $F_{k_0}(\mathbf{p}) > 0$ .

We prove only the first inequality, since the second is obtained by analogous reasoning. To this end, we increase all components of  $p_j$  ( $j \neq i_0$ ) if necessary, and thus achieve the price vector  $\mathbf{p}' = \langle p_{i_0}, p_{i_0}, \dots, p_{i_0} \rangle$ . Because of the property of strict gross substitutability,  $0 = F_{i_0}(\mathbf{p}') > F_{i_0}(\mathbf{p})$ , as was to be proved. The same argument also implies that the set  $\{\lambda \mathbf{p}^*\}$  ( $\lambda > 0$ ) spans the totality of equilibrium vectors.

We now continue with the proof of the theorem. Let  $\mathbf{p}(t)$  be the solution of (9.4.14) for an initial vector  $\mathbf{p}(0) \gg \mathbf{0}$ . We shall establish in (9.4.18) that  $\mathbf{p}(t) \gg \mathbf{0}$ , thus guaranteeing that  $\mathbf{F}[\mathbf{p}(t)]$  is meaningful for all  $t$ . For

the present we proceed as if this has been done. (The reader may check afterwards that the arguments are all consistent.) We now introduce the functions

$$\Lambda(t) = \max_{0 \leq j \leq r} p_j(t), \quad \lambda(t) = \min_{0 \leq j \leq r} p_j(t). \quad (9.4.15)$$

The definition implies the existence of an index  $i$  such that

$$\overline{\lim}_{h \rightarrow 0+} \frac{\Lambda(t+h) - \Lambda(t)}{h} \leq \overline{\lim}_{h \rightarrow 0+} \frac{p_i(t+h) - p_i(t)}{h} = \frac{dp_i(t)}{dt}$$

( $h \rightarrow 0+$  means that  $h$  approaches zero through positive values only) and

$$p_i(t) = \max_{0 \leq j \leq r} p_j(t).$$

Hence by (I),

$$\overline{\lim}_{h \rightarrow 0+} \frac{\Lambda(t+h) - \Lambda(t)}{h} \leq H_i(F_i[\mathbf{p}(t)]) \leq 0, \quad (9.4.16)$$

with strict inequality if  $\mathbf{p}(t)$  is not of the form  $\lambda \mathbf{p}^*$  for some positive  $\lambda$ . In a similar manner, we obtain

$$\overline{\lim}_{h \rightarrow 0+} \frac{\lambda(t+h) - \lambda(t)}{h} \geq 0, \quad (9.4.17)$$

with strict inequality except when  $\mathbf{p}(t)$  is of the form  $\lambda \mathbf{p}^*$  ( $\lambda > 0$ ). The inequalities (9.4.16) and (9.4.17) show that  $\Lambda(t)$  is nonincreasing and  $\lambda(t)$  is nondecreasing since each is continuous. In particular,

$$0 < \lambda(0) \leq \lambda(t) \leq \Lambda(t) \leq \Lambda(0); \quad (9.4.18)$$

hence the solution  $\mathbf{p}(t)$  is always bounded above and below, and each component of  $p_j(t)$  strictly avoids zero. Let

$$\tilde{\Lambda} = \lim_{t \rightarrow \infty} \Lambda(t),$$

and let  $\tilde{\mathbf{p}}$  be a limit point of the sequence  $\mathbf{p}(t)$ ; i.e.,

$$\tilde{\mathbf{p}} = \lim_{\mu \rightarrow \infty} \mathbf{p}(t_\mu)$$

for some sequence  $t_\mu \rightarrow \infty$ . Let  $\tilde{\mathbf{p}}(t)$  be the solution of (9.4.14) with initial vector  $\tilde{\mathbf{p}}$ , and define

$$\tilde{\Lambda}(t) = \max_{0 \leq j \leq r} \tilde{p}_j(t).$$

Then

$$\tilde{\Lambda}(t) = \lim_{\mu \rightarrow \infty} \Lambda(t + t_\mu) = \tilde{\Lambda},$$

which implies

$$\lim_{h \rightarrow 0} \frac{\tilde{\Lambda}(h) - \tilde{\Lambda}(0)}{h} = 0.$$

By (9.4.16),  $\tilde{\mathbf{p}}$  is a multiple of  $\mathbf{p}^*$  and hence an equilibrium vector. Since  $\tilde{\Lambda}$  is uniquely determined, we infer that  $\tilde{\mathbf{p}} = \tilde{\Lambda} \cdot \mathbf{p}^*$  is also uniquely determined. Therefore  $\mathbf{p}(t)$  itself converges to  $\tilde{\mathbf{p}}$ , and the proof of the theorem is complete.

Note that the proof of this theorem does not refer to the Walras law. By contrast, if weak rather than strict gross substitutability is assumed, i.e., if we require only that  $\partial F_i / \partial p_j \geq 0$  for  $i \neq j$  and  $\mathbf{p} \gg \mathbf{0}$ , the validity of the Walras law must be postulated.

► **THEOREM 9.4.4.** If the excess demand functions  $\mathbf{F}(\mathbf{p})$  are defined, differentiable for  $\mathbf{p} \gg \mathbf{0}$ , and homogeneous, and if they satisfy the Walras law and possess the property of weak gross substitutability, then the price adjustment process characterized by (9.4.14) is stable.

(As usual, we assume the existence of a strict positive equilibrium vector.)

*Proof.* Similarly to the proof of Theorem 9.4.3, the relation (9.4.18) holds for the present case; hence the solution  $\mathbf{p}(t)$  of the system (9.4.14) with any initial vector  $\mathbf{p}(0) \gg \mathbf{0}$  is bounded and positive.

Now consider the function

$$R(t) = \max_{i \in S(t)} \left\{ \frac{F_i(t)}{p_i(t)} \right\} \quad (9.4.19)$$

( $F_i(t)$  is an abbreviation for  $F_i[\mathbf{p}(t)]$ ), where

$$S(t) = \left\{ i \left| \frac{H_i[F_i(t)]}{p_i(t)} = \max_{0 \leq j \leq r} \frac{H_j[F_j(t)]}{p_j(t)} \right. \right\}.$$

It is noted that  $R(t) = 0$  or  $R(t) > 0$  according as  $\mathbf{p}(t)$  is or is not an equilibrium vector. For if  $R(t) \leq 0$ , it follows from the definition of  $R(t)$  that  $H_i(F_i[\mathbf{p}(t)]) \leq 0$ , where  $i = 0, 1, \dots, r$ ; and since  $H_i$  is sign-preserving,  $F_i[\mathbf{p}(t)] \leq 0$ . From these inequalities and the Walras law, it follows that  $F_i[\mathbf{p}(t)] = 0$  and  $\mathbf{p}(t)$  is an equilibrium vector.

Consider

$$W(t) = \overline{\lim}_{h \rightarrow 0} \frac{R(t+h) - R(t)}{h} \leq \frac{d}{dt} \left( \frac{F_i[\mathbf{p}(t)]}{p_i(t)} \right), \quad (9.4.20)$$

where the last inequality holds for an appropriate index  $i$  in  $S(t)$ . We obtain

$$\begin{aligned} \frac{d}{dt} \left( \frac{F_i(t)}{p_i(t)} \right) &= - \frac{F_i(t)}{p_i^2(t)} \frac{dp_i(t)}{dt} + \frac{1}{p_i(t)} \sum_{j=0}^r \frac{\partial F_i}{\partial p_j} \frac{dp_j}{dt} \\ &= - \frac{F_i(t) H_i[F_i(t)]}{p_i^2(t)} + \frac{1}{p_i(t)} \sum_{j=0}^r \frac{\partial F_i}{\partial p_j} H_j[F_j(t)]. \end{aligned}$$

Applying the property of weak gross substitutability, and taking account of the meaning of the index  $i$  and the homogeneity of  $\mathbf{F}$ , we obtain

$$\begin{aligned} \frac{1}{p_i(t)} \sum_{j=0}^r \frac{\partial F_i}{\partial p_j} H_j[F_j(t)] &= \frac{H_i[F_i(t)]}{p_i(t)} \frac{\partial F_i}{\partial p_i} + \frac{1}{p_i} \sum_{j \neq i} \frac{\partial F_i}{\partial p_j} H_j[F_j(t)] \\ &\leq \frac{H_i[F_i(t)]}{p_i(t)} \frac{\partial F_i}{\partial p_i} + \frac{H_i[F_i(t)]}{p_i^2(t)} \sum_{j \neq i} \frac{\partial F_i}{\partial p_j} p_j(t) \\ &= \frac{H_i[F_i(t)]}{p_i^2(t)} \sum_{j=0}^r \frac{\partial F_i}{\partial p_j} p_j(t) = 0. \end{aligned}$$

Hence the inequality (9.4.20) may be written as

$$W(t) \leq - \frac{F_i(t) H_i[F_i(t)]}{p_i^2(t)}. \quad (9.4.21)$$

Since  $H_i$  is sign-preserving, we infer from (9.4.21), together with the observation following (9.4.19), that

$$W(t) \leq 0, \quad (9.4.22)$$

with equality if and only if  $F[\mathbf{p}(t)] = 0$ .

Thus the function  $R(t)$  is nonincreasing and nonnegative (the latter by virtue of the Walras law); and hence  $R(t)$  converges to  $\tilde{R}$  as  $t$  tends to infinity. Now let  $\tilde{\mathbf{p}}$  be a limit point of  $\mathbf{p}(t_v)$  for some sequence  $t_v$  approaching infinity. Define  $\tilde{R}(t)$  and  $\tilde{W}(t)$  for the solution  $\tilde{\mathbf{p}}(t)$  with initial vector  $\tilde{\mathbf{p}}$  as  $R(t)$  and  $W(t)$  are defined for  $\mathbf{p}(t)$ . Then

$$\tilde{R}(t) = \lim_{v \rightarrow \infty} R(t + t_v) = \tilde{R},$$

and hence  $\tilde{W}(t) = 0$  at  $t = 0$ .

By (9.4.22),  $\tilde{\mathbf{p}}$  is an equilibrium price vector. We may now define for the solution  $\mathbf{p}(t)$ ,

$$\Lambda(t) = \max_j \frac{p_j(t)}{\tilde{p}_j} \quad \text{and} \quad \lambda(t) = \min_j \frac{p_j(t)}{\tilde{p}_j}. \quad (9.4.23)$$

Although the reasoning in Theorem 9.4.3 was carried out with respect to the equilibrium vector  $\mathbf{p}^*$ , it is readily shown that it remains correct for  $\Lambda(t)$  and  $\lambda(t)$  as defined in (9.4.23). Hence

$$\lim_{t \rightarrow \infty} \Lambda(t) \quad \text{and} \quad \lim_{t \rightarrow \infty} \lambda(t)$$

both exist. Moreover, since  $\lim \mathbf{p}(t_v) = \tilde{\mathbf{p}}$ , we have

$$\lim_{t \rightarrow \infty} \Lambda(t) = \lim_{t \rightarrow \infty} \lambda(t) = 1.$$

From this, in conjunction with (9.4.23), it follows that

$$\lim_{t \rightarrow \infty} \mathbf{p}(t) = \tilde{\mathbf{p}},$$

and the proof of the theorem is complete.

Valid adaptations of Theorems 9.4.1–4 can be formulated for the normalized price adjustment process.

**\*9.5. Global stability (continued).** In the last section we established global stability in one case by showing that the norm measured by the Euclidean distance between the solution and the equilibrium price vector decreased with time. We subsequently achieved sharper results by using other norms. In this section we again employ norms other than the Euclidean distance for the same purpose under different types of hypothesis. The three principal theorems of this section are formulated in terms of the normalized process. By suitable modifications, these theorems may be translated into propositions that are relevant for the nonnormalized process.

► **THEOREM 9.5.1.** The normalized process (9.2.8) is globally stable if there exists a set of positive constants  $\langle c_1, c_2, \dots, c_r \rangle$  such that for each  $j = 1, 2, \dots, r$

$$G_{jj} < 0, \quad c_j|G_{jj}| > \sum_{k \neq j} c_k|G_{kj}|, \quad (9.5.1)$$

where  $G_{kj} = \partial G_k / \partial q_j$ .

*Remark 9.5.1.* The condition of a dominant diagonal says that the price of the  $k$ th commodity influences the demand for that commodity more than it influences the demand for other commodities.

*Proof.* Consider the function

$$V(t) = \sum_{j=1}^r c_j|G_j|. \quad (9.5.2)$$

It is easily shown that

$$\lim_{h \rightarrow 0+} \frac{V(t+h) - V(t)}{h} = W(t)$$

exists. This value of  $W(t)$  may be calculated as

$$W(t) = \sum_j c_j \operatorname{sign} G_j(t) \sum_{k=1}^r \frac{\partial G_j}{\partial q_k} G_k + \sum_{j'} c_j \left| \sum_{k=1}^r \frac{\partial G_j}{\partial q_k} G_k \right|, \quad (9.5.3)$$

the first sum being extended over the indices  $j$  for which  $G_j(t) \neq 0$  and the second sum over the indices  $j$  for which  $G_j(t) = 0$ . The usual meaning is attached to  $\operatorname{sign}$ , namely

$$\operatorname{sign} a = \begin{cases} +1 & a > 0, \\ 0 & a = 0, \\ -1 & a < 0. \end{cases}$$

We shall prove that  $-W(t)$  is positive unless  $G_j = 0$  for all  $j$  (i.e., the solution is at an equilibrium price vector). We consider the coefficient  $L_{k_0}(t)$  in  $-W(t)$  of  $G_{k_0}(t)$ , which we suppose is not zero. Specifically,

$$\begin{aligned} L_{k_0}(t) &= -c_{k_0} \frac{\partial G_{k_0}}{\partial q_{k_0}} \operatorname{sign} G_{k_0}(t) - \sum_{j \neq k_0} c_j \operatorname{sign} G_j(t) \frac{\partial G_j}{\partial q_{k_0}} \\ &\quad - \sum_{j \neq k_0} c_j \operatorname{sign} \left( \sum_{k=1}^r \frac{\partial G_j}{\partial q_k} G_k \right) \cdot \frac{\partial G_j}{\partial q_{k_0}}. \end{aligned}$$

Hence

$$G_{k_0}(t) L_{k_0}(t) \geq c_{k_0} \left| \frac{\partial G_{k_0}}{\partial q_{k_0}} \right| G_{k_0}(t) - \left( \sum_{j \neq k_0} c_j \left| \frac{\partial G_j}{\partial q_{k_0}} \right| \right) |G_{k_0}(t)|, \quad (9.5.4)$$

and by the hypothesis the right-hand side is strictly positive. Adding the inequalities (9.5.4) for the different  $k_0$  (note that the terms in which  $G_{k_0}(t) = 0$  make no contribution to the value of  $W(t)$  and are thus of no relevance), we obtain  $W(t) < 0$  except when  $G_k(t) = 0$  for all  $k$ . The norm  $V(t)$ , as defined by (9.5.2), strictly decreases, and global stability results as in Theorem 9.3.4.

A dual result to Theorem 9.5.1 is as follows.

► **THEOREM 9.5.2.** The normalized process (9.2.8) is globally stable if there exists a set of positive constants  $\langle c_1, c_2, \dots, c_r \rangle$  such that for each  $j = 1, 2, \dots, r$

$$G_{jj} < 0, \quad c_j |G_{jj}| > \sum_{k \neq j} c_k |G_{jk}|. \quad (9.5.5)$$

*Remark 9.5.2.* The hypothesis may be loosely interpreted to mean that the price of the  $k$ th commodity is the major influence on the excess demand function of that commodity.

The difference between Theorem 9.5.1 and Theorem 9.5.2 is that (9.5.1) pertains to the column sums and (9.5.5) to the row sums of the matrix  $\|G_{jk}\|$ . These sets of conditions may be construed as dual criteria, each yielding global stability. The proofs likewise proceed in a dual fashion.

*Proof of Theorem 9.5.2.* Consider the norm function

$$V(t) = \max_j \frac{|G_j|}{c_j}. \quad (9.5.6)$$

From the point of view of linear space theory, (9.5.6) may be regarded as conjugate (dual) to (9.5.2). Here also we shall prove that  $dV/dt < 0$  except at equilibrium, wherever the derivative exists. Suppose that

$$V(t) = \frac{G_{j_0}[\mathbf{q}(t)]}{c_{j_0}} \operatorname{sign} G_{j_0}[\mathbf{q}(t)]$$

and  $G_{j_0}[\mathbf{q}(t)] \neq 0$ . Then

$$\begin{aligned} \frac{dV}{dt} &= \frac{\operatorname{sign} G_{j_0}[\mathbf{q}(t)]}{c_{j_0}} \sum_k \frac{\partial G_{j_0}}{\partial p_k} G_k \leq -\frac{|G_{j_0}|}{c_{j_0}} \left( -\frac{\partial G_{j_0}}{\partial p_{j_0}} \right) \\ &\quad + \frac{1}{c_{j_0}} \sum_{k \neq j_0} c_k \left| \frac{\partial G_{j_0}}{\partial p_k} \right| \frac{|G_k|}{c_k}. \end{aligned}$$

But by the definition of  $j_0$

$$\frac{|G_j|}{c_j} \leq \frac{|G_{j_0}|}{c_{j_0}},$$

and hence we have

$$\frac{dV}{dt} \leq -\frac{|G_{j_0}|}{c_{j_0}} \left| \frac{\partial G_{j_0}}{\partial p_{j_0}} \right| + \frac{|G_{j_0}|}{c_{j_0}^2} \sum_{k \neq j_0} c_k \left| \frac{\partial G_{j_0}}{\partial p_k} \right| < 0, \quad (9.5.7)$$

the last inequality following from the hypothesis of the theorem and the fact that  $|G_{j_0}(t)| \neq 0$ . Of course, when  $\mathbf{G} = \mathbf{0}$  we are at an equilibrium price vector and we stay there for all  $t$ . When  $dV/dt$  does not exist, we estimate

$$\lim_{h \rightarrow 0+} \frac{V(t+h) - V(t)}{h},$$

which always exists, and modify the arguments accordingly (see Theorem 9.5.1). Global stability follows from (9.5.7) in the usual way.

Another criterion of a different sort which also yields global stability is the requirement that  $\mathbf{B}(\mathbf{q}) = \|\partial G_i(\mathbf{q})/\partial q_j\|$  be negative quasi-definite.

(See p. 305 for the definition of a negative quasi-definite matrix.) A slight generalization of this criterion is embodied in the following theorem.

►THEOREM 9.5.3. Let  $\mathbf{S}$  denote a positive definite matrix of order  $r$ . If  $\mathbf{B}(\mathbf{q})'\mathbf{S} + \mathbf{S}\mathbf{B}(\mathbf{q})$  is negative definite for every price vector  $\mathbf{q}$ , then the normalized process is globally stable.

(Note that when  $\mathbf{S}$  is the identity matrix, the hypothesis reduces to the statement that  $\mathbf{B}(\mathbf{q})$  is negative quasi-definite.)

*Proof.* Here we introduce the metric

$$V(t) = (\mathbf{G}, \mathbf{S}\mathbf{G}) \quad (9.5.8)$$

for  $\mathbf{G}[\mathbf{q}(t)]$ . Differentiation of  $V(t)$  produces

$$V'(t) = (\mathbf{G}, [\mathbf{B}(\mathbf{q})'\mathbf{S} + \mathbf{S}\mathbf{B}(\mathbf{q})]\mathbf{G}),$$

which is strictly negative unless  $\mathbf{G} = \mathbf{0}$ . Global stability now follows in a routine fashion.

We close this section by outlining a general approach to the problem of ascertaining the stability of dynamic processes.

Consider the system of differential equations

$$\frac{dx_i}{dt} = f_i(\mathbf{x}) \quad (i = 1, \dots, r), \quad (9.5.9)$$

where  $\langle f_1(\mathbf{x}), \dots, f_r(\mathbf{x}) \rangle$  is a continuous vector function defined on a set  $X \subset E^r$ .

It is assumed that the set  $X$  is closed and has a nonvoid interior  $X^0$ ; and that the solution  $\mathbf{x}(t; \mathbf{x}^0)$  to the system with any initial position  $\mathbf{x}^0$  in  $X^0$  is uniquely determined, is continuous with respect to  $\mathbf{x}^0$ , and remains in  $X$ .

Let  $E$  be the set of all equilibrium vectors in  $X$ , i.e.,

$$E = \{\bar{\mathbf{x}} | \bar{\mathbf{x}} \text{ in } X, \mathbf{f}(\bar{\mathbf{x}}) = \mathbf{0}\},$$

and define the distance from a point  $\mathbf{x}$  in  $X$  to the set  $E$  by

$$V(\mathbf{x}) = \inf_{\bar{\mathbf{x}} \text{ in } E} |\mathbf{x} - \bar{\mathbf{x}}|^2.$$

The system (9.5.9) is called *quasi-stable* if, for any initial position  $\mathbf{x}^0$  in  $X^0$ ,

$$\lim_{t \rightarrow \infty} V[\mathbf{x}(t; \mathbf{x}^0)] = 0.$$

We state without proof the following general theorem.

►THEOREM 9.5.4. If for any initial position  $\mathbf{x}^0$  in  $X^0$ , the solution to the system (9.5.9) is bounded, and if there exists a function  $\phi(\mathbf{x})$  defined on  $X$

which satisfies a Lipschitz condition in  $X$  such that  $D_{\phi,f}(\mathbf{x}) < 0$  for any  $\mathbf{x}$  in  $X - E$ , then the system (9.5.9) is quasi-stable.

The function  $D_{\phi,f}$  is defined as follows: In general,

$$D_{\phi,f}(\mathbf{x}) = \overline{\lim}_{\Delta x \rightarrow 0} \sum \frac{\Delta_i \phi(\mathbf{x})}{\Delta x_i} f_i(\mathbf{x}),$$

where

$$\begin{aligned} \Delta_i \phi(\mathbf{x}) &= \phi(x_1 + \Delta x_1, \dots, x_i + \Delta x_i, x_{i+1}, \dots, x_r) \\ &\quad - \phi(x_1 + \Delta x_1, \dots, x_{i-1} + \Delta x_{i-1}, x_i, x_{i+1}, \dots, x_r). \end{aligned}$$

If  $\phi$  is differentiable, then

$$D_{\phi,f} = \sum_i \frac{\partial \phi}{\partial x_i} f_i.$$

The theorem can be proved by a suitable adaptation of the reasoning used in proving Theorems 9.4.3 and 9.4.4 of the previous section.

As a special application of this theorem, consider the situation in which  $\mathbf{f}(\mathbf{x})$  has a symmetrical matrix of the partial derivatives, i.e.,  $\partial f_i / \partial x_j = \partial f_j / \partial x_i$  for all  $i$  and  $j$ . From the theory of exact differentials, we infer the existence of a (potential) function  $\phi$  such that

$$-\frac{\partial \phi}{\partial x_i} = f_i(\mathbf{x}).$$

Therefore

$$\sum_{i=1}^r \frac{\partial \phi}{\partial x_i} f_i(\mathbf{x}) = - \sum_{i=1}^r f_i(\mathbf{x})^2 < 0, \quad \text{if } \mathbf{f}(\mathbf{x}) \neq \mathbf{0},$$

and the theorem may be applied. In this connection the reader may consult the notes at the close of the chapter (see also Problem 14).

In line with these concepts we give a general result on quasi-stability which applies in the case of an excess demand vector function induced by a utility maximization for a single consumer.

Let  $U(\mathbf{x})$  be a utility function defined for all  $\mathbf{x} > \mathbf{0}$  in  $E^{r+1}$  and obeying the inequality

$$\frac{\partial U}{\partial x_i} > 0 \tag{9.5.9}$$

for at least one index  $i$ . (This essentially means that if budget considerations are disregarded, then no consumption vector exists which will satiate the individual's desires.)

The individual consumer selects the consumption vector which maximizes his utility function subject to his budget constraint. Formally, we postulate the existence of a unique  $\mathbf{x}^*(\mathbf{p})$  for which

$$U(\mathbf{x}^*) = \max_{\mathbf{x}} U(\mathbf{x})$$

under the condition

$$\sum_{i=0}^r p_i x_i = \sum_{i=0}^r p_i x_i^0$$

and  $\mathbf{x} \geq \mathbf{0}$ . Here,  $\mathbf{p} > \mathbf{0}$  represents a fixed price vector and  $\mathbf{x}^0 \gg \mathbf{0}$  is interpreted as the fixed vector of initial holdings. The maximization can be carried out by the techniques of Lagrange multipliers. There result the necessary conditions

$$\frac{\partial U}{\partial x_i}(\mathbf{x}^*) \leq \lambda p_i, \quad (i = 0, 1, \dots, r), \quad (9.5.10)$$

with equality in the  $i$ th relation if  $x_i^* > 0$ . Assumption (9.5.9) implies  $\lambda > 0$ .

For this setup we define the price adjustment process by (9.4.14), where  $\mathbf{F}(\mathbf{p}) = \mathbf{x}^0 - \mathbf{x}^*(\mathbf{p})$ . We can now prove that this process is quasi-stable. To this end, consider the function  $\phi[\mathbf{p}(t)] = U(\mathbf{x}^*(\mathbf{p}))$ . We have

$$\phi[\mathbf{p}(t)] = \sum_{i=0}^r \frac{\partial U}{\partial x_i}(\mathbf{x}^*) \frac{dx_i^*}{dt} \leq \lambda \sum_{i=0}^r p_i \frac{dx_i^*}{dt}, \quad (9.5.11)$$

since either  $x_i^* > 0$  and equality holds in (9.5.10) or  $x_i^* = 0$  and then clearly  $\dot{x}_i^* \geq 0$ .

With the aid of (9.4.14) and the identity

$$\sum_{i=0}^r p_i \frac{\partial x_i^*}{\partial p_j} = -x_j^* + x_j^0 \quad (j = 0, 1, \dots, r)$$

resulting from differentiation of the budget constraint, we can reduce (9.5.11) to

$$\phi(t) \leq -\lambda \sum_{j=0}^r (x_j^0 - x_j^*) H_j(F_j(t)) \leq 0, \quad (9.5.12)$$

with strict inequality unless  $F_j(\mathbf{p}(t)) = 0$ . Quasi-stability is a consequence of the inequality (9.5.12) in the usual way.

For further stability results we refer the reader to Problems 5 through 15 and to a discussion of their solutions on page 355.

**\*9.6 A difference-equations formulation of global stability.** In this section we describe the price adjustment mechanism over time by means of a system of difference equations. Explicitly,

$$\mathbf{p}(t+1) = \max \{ \mathbf{0}, \mathbf{p}(t) + \rho \mathbf{F}[\mathbf{p}(t)] \} \quad (t = 0, 1, 2, \dots), \quad (9.6.1)$$

where  $\mathbf{p}(t)$  is the price vector at time  $t$  and  $\mathbf{F}$  represents the excess demand function and  $\rho$  is a fixed positive constant.

This formulation is harder to analyze than the differential-equations formulation. Nevertheless, it will permit us to prove the existence of an equilibrium price vector in a simpler manner. Furthermore, in the circumstance of strict gross substitutability, where the equilibrium vector is unique, we shall establish stability in the large.

We assume as always that the Walras law,

$$\sum_{i=0}^r p_i F_i(p_0, p_1, \dots, p_r) = 0, \quad (9.6.2)$$

is valid for any nonnegative price vector.

The system investigated here corresponds to the nonnormalized process, and in this case it is customary to suppose that the excess demand functions are homogeneous functions of degree zero.

First we shall simply prove the existence of an equilibrium price vector. Recall that an equilibrium price is a price vector  $\mathbf{p}^* = \langle p_0^*, p_1^*, \dots, p_r^* \rangle$  for which

$$F_i(\mathbf{p}^*) \leq 0. \quad (9.6.3)$$

Let every nonzero price vector be normalized as

$$\sum_{i=0}^r p_i = 1.$$

We denote the set of all normalized price vectors by  $\mathcal{Y}$ . We observe that for  $\mathbf{p} \in \mathcal{Y}$

$$\sum_{i=0}^r \max [0, p_i + \rho F_i(\mathbf{p})] = \lambda(\mathbf{p}) > 0.$$

Otherwise  $p_i + \rho F_i(\mathbf{p}) \leq 0$  for all  $i$ , and thus

$$\sum_{i=0}^r p_i^2 = \sum_{i=0}^r p_i^2 + \rho \sum_{i=0}^r p_i F_i(\mathbf{p}) \leq 0,$$

the first equality being true by virtue of the Walras law. But then  $p_i = 0$  for all  $i$ , which is contradictory to the supposition that  $\mathbf{p} \in \mathcal{Y}$ .

We now define a mapping  $\mathbf{T}$  of  $\mathcal{Y}$  into  $\mathcal{Y}$  by

$$\mathbf{T}(\mathbf{p}) = \frac{1}{\lambda(\mathbf{p})} \max [\mathbf{0}, \mathbf{p} + \rho \mathbf{F}(\mathbf{p})]$$

(the maximum operation is taken componentwise). It is obvious that  $\mathbf{T}(\mathbf{p})$  is a continuous mapping of  $\mathcal{Y}$  into  $\mathcal{Y}$ . We may therefore apply the Brouwer fixed-point theorem and infer the existence of  $\bar{\mathbf{p}} \in \mathcal{Y}$  such that  $\bar{\mathbf{p}} = \mathbf{T}(\bar{\mathbf{p}})$ , that is,

$$\lambda(\bar{\mathbf{p}})\bar{\mathbf{p}} = \max [\mathbf{0}, \bar{\mathbf{p}} + \rho \mathbf{F}(\bar{\mathbf{p}})].$$

Then

$$\lambda(\bar{\mathbf{p}})\bar{p}_i \geq \bar{p}_i + \rho F_i(\bar{\mathbf{p}}) \quad (i = 0, 1, \dots, r), \quad (9.6.4)$$

with equality for all  $i$  such that  $\bar{p}_i > 0$ . Hence, multiplying by  $\bar{p}_i$  and adding the resulting equations,

$$\lambda(\bar{\mathbf{p}}) \sum_{i=0}^r \bar{p}_i^2 = \sum_{i=0}^r \bar{p}_i^2 + \rho \sum_{i=0}^r \bar{p}_i F_i(\bar{\mathbf{p}}) = \sum_{i=0}^r \bar{p}_i^2.$$

Therefore  $\lambda(\bar{\mathbf{p}}) = 1$ , and (9.6.4) shows that  $\bar{\mathbf{p}}$  is an equilibrium vector.

The existence of equilibrium having been established, we turn to a discussion of the dynamics of stability. The following theorem gives one criterion for stability of the discrete time system (9.6.1).

► **THEOREM 9.6.1.** Let  $\mathbf{p}^*$  denote a strictly positive equilibrium vector, and suppose that the excess demand vector function  $\mathbf{F}$  satisfies the property of strict gross substitutability in the large. Then there exists a positive  $\rho_0$  such that for  $\rho < \rho_0$  the price system  $\mathbf{p}(t)$  satisfying (9.6.1) converges to an equilibrium price vector.

*Remark 9.6.1.* The conclusion of this theorem can be interpreted as follows: If only relative prices are relevant, then the discrete-time price adjustment process, as defined by the system (9.6.1) is, globally stable.

*Proof.* Consider the norm function

$$V(t) = \sum_{i=0}^r [p_i(t) - p_i^*]^2.$$

A direct calculation, using (9.6.1), shows that

$$\begin{aligned} V(t) - V(t+1) &= \sum_{i=0}^r p_i^2(t) - p_i^2(t+1) - 2[p_i(t) - p_i(t+1)]p_i^* \\ &\geq \sum_{i=0}^r p_i^2(t) - p_i^2(t+1) + 2\rho \sum_{i=0}^r p_i^* F_i[\mathbf{p}(t)]. \end{aligned} \quad (9.6.5)$$

Let  $R$  denote these indices of the set  $\{1, 2, \dots, r\}$  for which  $p_i(t+1) = p_i(t) + \rho F_i[\mathbf{p}(t)]$ , and  $R'$  the complementary set of indices, those for which  $p_i(t+1) = 0$ . We note that

$$\begin{aligned}\sum_{i \in R} [p_i^2(t) - p_i^2(t+1)] &= -\rho^2 \sum_{i \in R} F_i^2[\mathbf{p}(t)] - 2\rho \sum_{i \in R} p_i(t) F_i[\mathbf{p}(t)] \\ &= -\rho^2 \sum_{i \in R} F_i^2[\mathbf{p}(t)] + 2\rho \sum_{i \in R'} p_i(t) F_i[\mathbf{p}(t)],\end{aligned}$$

the last equality resulting by virtue of the Walras law (9.6.2). Thus

$$\begin{aligned}\sum_{i=1}^r [p_i^2(t) - p_i^2(t+1)] &= -\rho^2 \sum_{i=1}^r F_i^2[\mathbf{p}(t)] + \sum_{i \in R'} \{[p_i(t)] + \rho F_i[\mathbf{p}(t)]\}^2 \\ &\geq -\rho^2 \sum_{i=1}^r F_i^2[\mathbf{p}(t)].\end{aligned}$$

We choose  $\rho$  satisfying

$$0 < \rho < \min_{p \neq \lambda p^*} \frac{2 \sum_{i=0}^r p_i^* F_i(\mathbf{p})}{\sum_{i=0}^r F_i^2(\mathbf{p})}. \quad (9.6.6)$$

That this is feasible is proved as follows. We expand the numerators and denominators of the right-hand side of (9.6.6) in the neighborhood of  $\mathbf{p}^*$ , and simplify to

$$S(\mathbf{p}) = \frac{\sum_{j,k=0}^r p_j \left( \frac{\partial F_j(\mathbf{p}^*)}{\partial p_k} + \frac{\partial F_k(\mathbf{p}^*)}{\partial p_j} \right) p_k}{\sum_{j,k=0}^r p_k p_j \sum_{i=0}^r \frac{\partial F_i(\mathbf{p}^*)}{\partial p_j} \frac{\partial F_i(\mathbf{p}^*)}{\partial p_k}}$$

[see (9.3.6)]. A further expansion of the two quadratic forms occurring in the numerator and denominator of  $S(\mathbf{p})$  will show that since strict gross substitutability prevails,

$$\inf_{p \neq \lambda p^*} S(\mathbf{p}) > 0.$$

We next observe by virtue of Theorem 9.4.2 that

$$\sum_{i=0}^r p_i^* F_i(\mathbf{p}) > 0, \quad \sum_{i=0}^r F_i^2(\mathbf{p}) > 0$$

for all  $\mathbf{p} \neq \lambda \mathbf{p}^*$ .

The continuity and homogeneity of  $F_i(\mathbf{p})$ , plus the fact that  $S(\mathbf{p})$  is bounded away from zero, enable us to deduce the validity of (9.6.6). For any selection of  $\rho$  satisfying (9.6.6), it follows from (9.6.5) that

$$V(t) > V(t+1) \quad (9.6.7)$$

unless  $\mathbf{p}(t) = \lambda \mathbf{p}^*$ . Consequently,

$$\lim_{t \rightarrow \infty} V(t)$$

exists and any limit point of  $\mathbf{p}(t)$  is a multiple of  $\mathbf{p}^*$ , as was shown in the proof of Theorem 9.3.4. Since it is now easy to establish the convergence of  $\mathbf{p}(t)$ , we omit the details; the proof of the theorem is hereby complete.

**9.7 Stability and expectations (Model I).** The previous models involved the adjustment of prices over time in response to changes in the excess demand functions, which are functions of the current market prices only. In this section and the next we examine the influence of anticipated future prices in addition to actual prices on the dynamic stability of the market system.

With this end in view, we propose to write the general dynamic equilibrium system as

$$\frac{d\mathbf{p}}{dt} = \mathbf{K}\mathbf{F}, \quad (9.7.1)$$

where  $\mathbf{p}$  is the price vector of the commodity bundle of  $r$  components,  $\mathbf{K}$  is an  $r \times r$  matrix of constants which determines how rapidly prices adjust to excess demands, and  $\mathbf{F}$  designates, as usual, the excess demand vector. Here, in general, each component of  $\mathbf{F}$  is a function of the current price vector  $\mathbf{p}$  and the expected future price vector  $\mathbf{p}^f$ . In discrete time the price vector  $\mathbf{p}^f$  refers to the anticipated prices for the next time period. In continuous time  $\mathbf{p}^f$  corresponds to anticipated prices in the next infinitesimal time period.

In our previous discussions we implicitly assumed that  $\mathbf{p}^f = \mathbf{p}$ , i.e., that expectations are static, and we accordingly took the matrix  $\mathbf{K}$  as in essence a diagonal matrix with positive components. Here, by contrast, we permit the expected future prices to deviate from current prices according to a determinative law and in this way influence the fluctuation of current prices. We shall show later that a stable dynamic system can absorb the effects of some extrapolation of price movements, together with the corresponding expected future price changes, and remain stable. The effects of expected future prices will be considered only with regard to the structure of local stability.

At equilibrium  $\mathbf{p} = \mathbf{p}^*$ , it is customary to suppose that  $\mathbf{p}^f = \mathbf{p}^*$ , i.e., that the expectations and the actual prices agree. Hence, by a Taylor expansion we may approximate  $F_i(\mathbf{p}, \mathbf{p}^f) = F_i(p_1, \dots, p_r; p_1^f, \dots, p_r^f)$  about an equilibrium price vector  $\mathbf{p}^*$  by the linear expression

$$F_i = \sum_{j=1}^r a_{ij}(p_j - p_j^*) + \sum_{j=1}^r b_{ij}(p_j^f - p_j^*), \quad (9.7.2)$$

which is valid in the neighborhood of  $\mathbf{p}^*$ .

The problem of ensuring the local stability of the system (9.7.1) can thus be reduced to the problem of stability of the linear system of differential equations

$$\frac{dz}{dt} = \mathbf{K}(\mathbf{A}z + \mathbf{B}w), \quad (9.7.3)$$

where  $z(t) = \mathbf{p}(t) - \mathbf{p}^*$  and  $w(t) = \mathbf{p}'(t) - \mathbf{p}^*$ . (The proof of this assertion is essentially the same as the argument of p. 306.)

The linear system (9.7.3) involves the two functions  $z(t)$  and  $w(t)$ . Therefore, to completely specify the dynamics of the price system it is necessary to indicate how  $w(t)$  is related to current prices. The simplest assumption is that changes in the expected future price of the  $i$ th commodity are induced by changes in its actual price according to the relationship

$$p_i' = p_i + \eta_i \frac{dp_i}{dt}, \quad (9.7.4)$$

where  $\eta_i$  is a constant. When  $\eta_i = 0$ , current prices are expected to persist. When  $\eta_i > 0$ , some multiple of the change in prices is added to the current price in arriving at the expected price; hence the expected price may be described as extrapolative. When  $\eta_i < 0$ , expected prices do not change as much as actual prices.

In vector notation, (9.7.4) may be written as

$$\mathbf{p}' = \mathbf{p} + \boldsymbol{\eta} \frac{d\mathbf{p}}{dt}, \quad (9.7.5)$$

where  $\boldsymbol{\eta}$  is a diagonal matrix. We generalize immediately and allow  $\boldsymbol{\eta}$  to be an arbitrary constant matrix. In the notation of (9.7.3),

$$w(t) = z(t) + \boldsymbol{\eta} \frac{dz(t)}{dt}. \quad (9.7.6)$$

Inserting (9.7.6) into (9.7.3), we obtain

$$\frac{dz}{dt} = [\mathbf{I} - \mathbf{K}\mathbf{B}\boldsymbol{\eta}]^{-1}\mathbf{K}(\mathbf{A} + \mathbf{B})z, \quad (9.7.7)$$

provided that the matrix  $(\mathbf{I} - \mathbf{K}\mathbf{B}\boldsymbol{\eta})^{-1}$  exists. We shall assume for convenience in what follows that the inverse matrix is well defined. In the case of static expectations ( $\mathbf{p}' = \mathbf{p}$ ), the corresponding dynamic system is

$$\frac{dz}{dt} = \mathbf{K}(\mathbf{A} + \mathbf{B})z. \quad (9.7.8)$$

We shall now develop conditions on the matrices appearing in the systems (9.7.7) and (9.7.8) that imply the stability or instability of the two systems together.

The problem discussed above may be stated formally as follows:

Let  $\mathbf{T}$  denote a stable matrix (all eigenvalues of  $\mathbf{T}$  have negative real parts). Two questions are of interest. First, under what conditions on the matrix  $\mathbf{Q}$  is  $\mathbf{QT}$  again stable? Second, what properties must  $\mathbf{T}$  and  $\mathbf{Q}$  have for  $\mathbf{T}$  and  $\mathbf{QT}$  to be simultaneously stable? Theorems 9.7.1 and 9.7.2 provide partial answers to these questions (see also Problems 1–4).

► **THEOREM 9.7.1.** Let  $\mathbf{T}$  be a stable  $M$ -matrix (see p. 255), and let  $\mathbf{Q}$  be a real diagonal matrix. Then  $\mathbf{QT}$  is stable if and only if the diagonal elements of  $\mathbf{Q}$  are strictly positive.

*Proof.* (Sufficiency.) The hypothesis asserts that the diagonal elements of  $\mathbf{Q}$  are strictly positive. Clearly  $\mathbf{QT}$  is an  $M$ -matrix. Moreover,  $-(\mathbf{QT})^{-1} = -\mathbf{T}^{-1}\mathbf{Q}^{-1}$  is positivity-preserving (by property (c) of p. 256), and hence  $\mathbf{QT}$  is stable.

The proof of necessity turns upon the following lemma.

► **LEMMA 9.7.1.** Let  $\mathbf{T}$  be a stable  $M$ -matrix of order  $r$ . If  $\mathbf{D}$  is a nonsingular diagonal matrix, then  $\mathbf{DT}$  has no eigenvalues on the imaginary axis.

*Proof.* Suppose to the contrary that  $\omega\lambda$  is an eigenvalue of  $\mathbf{Q} = \mathbf{DT}$ . (Here  $\omega^2 = -1$ , and obviously  $\lambda \neq 0$  since  $\mathbf{T}$  is nonsingular.) Consequently, there exists a nonzero vector  $\mathbf{x}$  such that

$$d_{ii} \sum_{j=1}^r t_{ij}x_j = \omega\lambda x_i \quad (i = 1, 2, \dots, r). \quad (9.7.9)$$

Adding a sufficiently large positive multiple  $\mu$  of the identity makes it certain that  $\mathbf{T} + \mu\mathbf{I} = \mathbf{S}$  is a nonnegative matrix. Equation (9.7.9) becomes

$$\sum_{j=1}^r S_{ij}x_j = \left( \mu + \frac{\omega\lambda}{d_{ii}} \right) x_i. \quad (9.7.10)$$

Taking absolute values and identifying  $y_i = |x_i|$ , we obtain

$$\mathbf{S}\mathbf{y} \geq \left( \min_i \left| \mu + \frac{\omega\lambda}{d_{ii}} \right| \right) \mathbf{y} \geq \mu\mathbf{y}$$

and  $\mathbf{y} > \mathbf{0}$ . From the characterization of the spectral radius of nonnegative matrices as given in Corollary 8.2.1 it follows that there exists an eigenvalue  $\lambda_0(\mathbf{S}) \geq \mu$ . Consequently  $\mathbf{T}$  has a nonnegative eigenvalue, contrary to the hypothesis.

*Proof of Theorem 9.7.1.* (Necessity.) We first observe that since  $\mathbf{QT}$  is stable,  $\mathbf{Q}$  is nonsingular.

Let  $\theta$  represent a real parameter varying between 0 and 1, and consider the family of matrices

$$\mathbf{U}(\theta) = \mathbf{Q}[\theta\mathbf{T} - (1 - \theta)\mathbf{I}].$$

For each  $\theta$  ( $0 \leq \theta \leq 1$ ), the bracketed term is clearly a stable  $M$ -matrix; therefore, by the lemma,  $\mathbf{U}(\theta)$  never has an imaginary eigenvalue. But the eigenvalues vary continuously, and hence for each  $\theta$  ( $0 \leq \theta \leq 1$ ) they always lie interior to the left half complex plane, since that is the situation for  $\theta = 1$ . Using this fact for  $\theta = 0$ , we infer that the diagonal elements of  $\mathbf{Q}$  are all positive, as was to be proved.

► **THEOREM 9.7.2.** If  $\mathbf{S}$  is symmetric and  $\mathbf{T}$  is a real negative quasi-definite matrix,  $\mathbf{ST}$  is stable if and only if  $\mathbf{S}$  is positive definite.

*Remark 9.7.1.* We remind the reader that a matrix  $\mathbf{T}$  is negative quasi-definite if and only if  $(\mathbf{x}, \mathbf{Tx})$  is negative for every nonzero real vector  $\mathbf{x}$ .

A real negative quasi-definite matrix is stable. We prove this assertion as follows. Suppose that

$$\mathbf{Tx} = \lambda \mathbf{x}, \quad (9.7.11)$$

where  $\mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_2$  and  $\lambda = \lambda_1 + i\lambda_2$ , with  $\mathbf{x}_1$  and  $\mathbf{x}_2$  real vectors not both zero. Rewriting the real and imaginary parts of (9.7.11), we obtain

$$\mathbf{Tx}_1 = \lambda_1 \mathbf{x}_1 - \lambda_2 \mathbf{x}_2, \quad \mathbf{Tx}_2 = \lambda_1 \mathbf{x}_2 + \lambda_2 \mathbf{x}_1.$$

Hence

$$(\mathbf{x}_1, \mathbf{Tx}_1) + (\mathbf{x}_2, \mathbf{Tx}_2) = \lambda_1[(\mathbf{x}_1, \mathbf{x}_1) + (\mathbf{x}_2, \mathbf{x}_2)],$$

and we conclude that  $\lambda_1 < 0$ .

*Proof of Theorem 9.7.2.* (Sufficiency.) If  $\mathbf{S}$  is positive definite,  $\mathbf{S}^{1/2}\mathbf{TS}^{1/2}$  is real negative quasi-definite and hence stable, where  $\mathbf{S}^{1/2}$  is real and positive definite. But  $\mathbf{RS}^{1/2}\mathbf{TS}^{1/2}\mathbf{R}^{-1} = \mathbf{ST}$  is stable for  $\mathbf{R} = \mathbf{S}^{1/2}$ , since it arises from a stable matrix by a similarity transformation.

(Necessity.) We show that if  $\mathbf{S}$  is nonsingular symmetric and  $\mathbf{T}$  is real negative quasi-definite, then  $\mathbf{ST}$  has no purely imaginary eigenvalue. To this end, suppose that there exists a nonnull vector  $\mathbf{x}$  such that  $\mathbf{STx} = \lambda \mathbf{x}$ . Premultiply both sides by  $\bar{\mathbf{x}}\mathbf{S}^{-1}$ , where  $\bar{\mathbf{x}}$  denotes the conjugate of  $\mathbf{x}$ :

$$(\bar{\mathbf{x}}\mathbf{S}^{-1}, \mathbf{STx}) = (\bar{\mathbf{x}}, \mathbf{Tx}) = \lambda(\bar{\mathbf{x}}, \mathbf{S}^{-1}\mathbf{x}). \quad (9.7.12)$$

Write  $\mathbf{x} = \mathbf{y} + i\mathbf{z}$  and  $\lambda = \alpha + i\beta$ , where  $\mathbf{y}$ ,  $\mathbf{z}$ ,  $\alpha$ , and  $\beta$  are all real. Equating real parts of (9.7.12) gives

$$(\mathbf{y}, \mathbf{Ty}) + (\mathbf{z}, \mathbf{Tz}) = \alpha[(\mathbf{y}, \mathbf{S}^{-1}\mathbf{y}) + (\mathbf{z}, \mathbf{S}^{-1}\mathbf{z})]. \quad (9.7.13)$$

The left-hand side is strictly negative, and hence  $\alpha \neq 0$ . The remainder of the proof follows the proof of Theorem 9.7.1. We omit the details.

The economic significance of Theorem 9.7.1 is plain. The relation of  $M$ -matrices and gross substitutes has been duly discussed. The fact that

the matrix  $\mathbf{T} = \mathbf{A} + \mathbf{B}$  has nonnegative off-diagonal elements means that the system contains no complementary commodities. If in addition the matrices  $\mathbf{K}$ ,  $\mathbf{B}$ , and  $\boldsymbol{\eta}$  are all diagonal, then Theorem 9.7.1 asserts that the expectationless system is stable if and only if all the  $k_{ii}$  are positive. Moreover, for the same  $\mathbf{T} = \mathbf{A} + \mathbf{B}$ , stability remains unaffected by the introduction of the expectational coefficients if and only if  $1 > k_{ii}b_{ii}\eta_{ii}$  for all  $i$ . Theorem 9.7.2 deals with another case, in which the income effects are essentially symmetric. Here we no longer require that  $\mathbf{K}$ ,  $\mathbf{B}$ , and  $\boldsymbol{\eta}$  be diagonal.

Further consequences along the lines of these theorems can be obtained by applying similarity transformations to the matrices  $\mathbf{A}$  and  $\mathbf{T}$  (see Problem 1).

**9.8 Stability and expectations (Model II).** A relationship between expected future prices  $\mathbf{p}^f$  and actual prices was expressed by (9.7.4) and more generally by (9.7.5).

Another such relationship of some economic merit is

$$\frac{dp_i^f}{dt} = \beta_i[p_i(t) - p_i^f(t)] \quad (i = 1, 2, \dots, r). \quad (9.8.1)$$

When  $\beta_i = 0$ , changes in current prices have no effect on expected prices. The condition for static expectations is now  $\beta_i = +\infty$ .

For a given time path of prices  $p_i(t)$ , equation (9.8.1) has the solution

$$p_i^f(t) = p_i^f(0)e^{-\beta_i t} + e^{-\beta_i t} \int_0^t \beta_i p_i(u)e^{\beta_i u} du, \quad (9.8.2)$$

where  $p_i^f(0)$  is the initial value of the expected price of the  $i$ th commodity. If the origin is in the sufficiently distant past, the term in (9.8.2) involving  $p_i^f(0)$  is negligible; hence  $p_i^f(t)$  may be taken as an exponentially weighted average of past prices. Consequently, (9.8.1) indicates that the expected price as formed is a properly weighted average of actual past prices [exponential in the case of (9.8.2)].

Still another rationale, attributable to Hicks, is as follows. If we think of time as a discontinuous variable, then the notion of "elasticity of expectations" may be formulated as

$$\frac{p_i^f(t) - p_i^f(t-1)}{p_i(t-1) - p_i^f(t-1)} = \beta_i, \quad (9.8.3)$$

where the prices are expressed in logarithms and  $\beta_i$  is a constant. When  $\beta_i = 0$ , changes in actual prices have no effect on expected price; when  $\beta_i = 1$ , current price is projected forward as people's expectations of the level of future prices. The value of  $\beta_i$  can be considered as a measure of the

elasticity of people's expectations between the extreme values of  $\beta = 0$  and  $\beta = 1$ .

If we drop the assumption that prices are expressed in logarithms, the differential-equations analog of (9.8.3) is (9.8.1).

We turn now to a discussion of the local stability of (9.7.1) in terms of the relationship (9.8.1). We seek to compare the conditions of stability for the expectationless system

$$\frac{dz}{dt} = \mathbf{K}(\mathbf{A} + \mathbf{B})z, \quad (9.8.4)$$

where  $\mathbf{K}$  is a diagonal matrix and  $\mathbf{A}$  and  $\mathbf{B}$  are as in (9.7.2), with the conditions of stability for the dynamic system with expected future prices accounted for. This system is represented by

$$\frac{dz}{dt} = \mathbf{K}(\mathbf{A}z + \mathbf{B}w), \quad \frac{dw}{dt} = \beta z - \beta w, \quad (9.8.5)$$

where  $z = \mathbf{p}(t) - \mathbf{p}^*$ ,  $w = \mathbf{p}'(t) - \mathbf{p}^*$ , and  $\beta$  is a diagonal matrix with the elements  $\beta_i$  along the diagonal and 0 elsewhere. The coefficient matrix of the system (9.8.5) is of the form

$$\mathbf{C} = \begin{pmatrix} \mathbf{KA} & \mathbf{KB} \\ \beta & -\beta \end{pmatrix},$$

where  $\tilde{\mathbf{A}} = \mathbf{KA}$  is the matrix  $\|k_{ii}a_{ij}\|$  and  $\tilde{\mathbf{B}} = \mathbf{KB}$  is the matrix  $\|k_{ii}b_{ij}\|$ . The coefficient matrix for the system (9.8.4) is of the form

$$\mathbf{D} = \mathbf{K}(\mathbf{A} + \mathbf{B}) = \tilde{\mathbf{A}} + \tilde{\mathbf{B}}.$$

In order to simplify matters, we assume that all present commodities are gross substitutes for one another and also that each future commodity is a gross substitute for each present commodity. That is, we assume that  $a_{ij} \geq 0$  ( $i \neq j$ ) and  $b_{ij} \geq 0$  ( $i, j = 1, \dots, r$ ). Since  $\mathbf{K}$  and  $\beta$  are diagonal nonnegative matrices, we see that  $\mathbf{C}$  and  $\mathbf{D}$  are both  $M$ -matrices.

► **THEOREM 9.8.1.** The system (9.8.5) with matrix  $\mathbf{C}$  is stable if and only if the system (9.8.4) with matrix  $\mathbf{D}$  is stable.

*Remark 9.8.1.* The principal argument employed in the proof of the theorem hinges on two observations indicated earlier: (a) an  $M$ -matrix  $\mathbf{T}$  is stable if and only if the spectral radius of the positive matrix  $\mathbf{S} = (\mathbf{T} + \mu\mathbf{I})$  is smaller than  $\mu$ , where  $\mu$  is positive and sufficiently large to make  $\mathbf{T} + \mu\mathbf{I}$  a nonnegative matrix (see p. 255); (b) the spectral radius  $\lambda_0(\mathbf{S})$  of a nonnegative matrix  $\mathbf{S}$  is characterized as the supremum of all  $\lambda$  for which there exists a positive vector  $\mathbf{x}$  satisfying  $\mathbf{Sx} \geq \lambda\mathbf{x}$ .

*Proof of Theorem 9.8.1.* Let  $\mathbf{U} = \mathbf{C} + \mu \mathbf{I}_{2r}$  and  $\mathbf{V} = \mathbf{D} + \mu \mathbf{I}_r$ , where  $\mu$  is large enough so that  $\mathbf{U}$  and  $\mathbf{V}$  are both nonnegative matrices. (Of course,  $\mathbf{U}$  is a matrix of order  $2r$  and  $\mathbf{V}$  is a matrix of order  $r$ .) Let  $\lambda_0(\mathbf{U})$  and  $\lambda_0(\mathbf{V})$  denote the spectral radii of  $\mathbf{U}$  and  $\mathbf{V}$ , respectively. Suppose that  $\mu$  is such that  $\mathbf{V}\mathbf{x} \geq \mu\mathbf{x}$ , where  $\mathbf{x} = \langle x_1, \dots, x_r \rangle$  and  $\mathbf{x} > \mathbf{0}$ , which is certainly the case if  $\mathbf{D}$  is not zero. Define the vector  $\mathbf{y}$  of  $2r$  coordinates by  $\mathbf{y} = \langle x_1, \dots, x_r; x_1, \dots, x_r \rangle$ . Then, by immediate check, we have  $\mathbf{U}\mathbf{y} \geq \mu\mathbf{y}$ , and hence

$$\lambda_0(\mathbf{U}) \geq \lambda_0(\mathbf{V}) \geq \mu. \quad (9.8.6)$$

Thus, if  $\mathbf{C}$  is stable we must have  $\mathbf{D}$  stable, since in the contrary event we obtain  $\lambda_0(\mathbf{U}) \geq \mu$ , in contradiction to the assumption that  $\mathbf{C}$  is stable. Next, consider  $\mu$  such that

$$\mathbf{U}\mathbf{y}' \geq \mu\mathbf{y}' \quad (9.8.7)$$

for some  $\mathbf{y}' > \mathbf{0}$ . The last  $r$  inequalities of (9.8.7) reduce to

$$\beta_i(y'_i - y'_{i+r}) \geq 0,$$

and hence  $y'_i \geq y'_{i+r} \geq 0$ . Define  $\mathbf{x}' = \langle y'_1, \dots, y'_r \rangle$ . Since  $\mathbf{U}$  is non-negative, we may increase the values of the last  $r$  components to the values  $y'_i$ , and the first  $r$  inequalities of (9.8.7) become

$$\mathbf{V}\mathbf{x}' = \tilde{\mathbf{A}}\mathbf{x}' + \tilde{\mathbf{B}}\mathbf{x}' + \mu\mathbf{I}\mathbf{x}' \geq \mathbf{U}\mathbf{y}' \geq \mu\mathbf{x}'. \quad (9.8.8)$$

Clearly,  $\mathbf{x}' > \mathbf{0}$ , and therefore

$$\lambda_0(\mathbf{V}) \geq \lambda_0(\mathbf{U}) \geq \mu. \quad (9.8.9)$$

Using the same reasoning as in the analysis of (9.8.6) we may infer from (9.8.9) the assertion that if  $\mathbf{D}$  is stable, then  $\mathbf{C}$  is stable. The proof is now complete.

It can be argued that the stability of the dynamic system under static expectations is plausible; hence, by the theorem proved, stability in general is plausible no matter what the inertia of the system or the elasticities of expectations.

**9.9 The von Neumann model of an expanding economy.** As we have seen, the equilibrium relations for the laws of supply and demand reduce mathematically to certain systems of equations. The problem is then to establish the existence of economically meaningful solutions to these equations, and it is solved by imposing various kinds of assumptions on the nature of the supply and demand functions. All theory of this sort—i.e., the sort discussed in Chapter 8—is concerned with the statics of the economic process; it yields no insight into the dynamic workings of the economy.

In this chapter (Sections 9.2–9.7) we have formulated the dynamic price adjustment process and investigated the stability of the system with respect to convergence of the price vector system to equilibrium. The price adjustment process can be regarded as an aspect of the short-run dynamics of the economy, in the sense that price changes upsetting the equilibrium situation are followed by readjustments toward equilibrium in response to the natural fluctuation of supply and demand.

We can also view the dynamics of the economy from a long-run perspective, in the sense of keeping account of actual amounts produced and consumed in each period, where the economy is assumed to operate in an efficient manner for an infinite time. A natural growth of production and consumption capacity appears to be associated with the economic process; we seek to determine the rate of this growth. This expansion process is known as the phenomenon of balanced growth.

We begin with a description of the classical model of an expanding economy proposed by von Neumann. Consider an economic system engaged in the production of  $n$  commodities ( $i = 1, \dots, n$ ). Let  $j = 1, \dots, m$  denote the activities that are carried out in the system. If the  $j$ th activity is carried out at the unit level of intensity, then  $a_{ij}$  and  $b_{ij}$  denote respectively the input and the output of the  $i$ th commodity in the  $j$ th activity. We make the natural assumption that  $a_{ij} \geq 0$ ,  $b_{ij} \geq 0$ . Let  $\mathbf{A} = \|a_{ij}\|$ ,  $\mathbf{B} = \|b_{ij}\|$ .

The level of intensity of the  $j$ th activity at time  $t$  will be denoted by  $y_j(t)$ . The price of the  $i$ th commodity will be denoted by  $p_i(t)$ , and the total amount of the  $i$ th commodity that is available by  $x_i(t)$ .

The basic assumptions imposed on the economic system are as follows:

$$(a) \quad x_i(t) \geq \sum_j a_{ij}y_j(t) \quad \text{for all } i \text{ and } t.$$

This is the obvious restriction that the amount of any commodity used at a particular time must not exceed the amount of that commodity available at that time.

$$(b) \quad x_i(t+1) = x_i(t) + \sum_j (b_{ij} - a_{ij})y_j(t) \quad \text{for all } i \text{ and } t.$$

This difference equation is the dynamic equation of the system. A continuous-time formulation of the dynamics of the system would involve replacing (b) by an analogous differential equation, but the essential features of the analysis would remain unchanged. There are, of course, prescribed initial conditions.

$$(c) \quad y_j(t) = y(t)y_j, \quad \text{where } y(t) = e^{\lambda t}.$$

The proportions between the intensities of the activities remain the same; only the total output increases with time. This assumption can be justified as an equilibrium condition for the economy.

$$(d) \sum_{i,j} p_i(t+1)b_{ij}y_j(t) - \sum_{i,j} p_i(t)a_{ij}y_j(t) \leq 0 \quad \text{for all } y_j(t).$$

There is no positive return from the system; it is a closed system.

$$(e) p_i(t) = p(t)p_i, \quad \text{where } p(t) = e^{-\mu t}.$$

The  $\lambda$  in assumption (c) is the expansion rate, and the  $\mu$  here is the discount rate of the system. Alternatively,  $e^{\mu t}$  can be regarded as the rate at which the interest of a unit of money accumulates with time.

By some manipulations of the assumptions we now develop various relations satisfied by the expansion rate and interest rate. If we substitute the values of  $p_i(t)$  and  $y_j(t)$  in (d), we obtain

$$\sum_{i,j} p_i e^{-\mu(t+1)} b_{ij} y_j e^{\lambda t} - \sum_{i,j} p_i e^{-\mu t} a_{ij} y_j e^{\lambda t} \leq 0,$$

or

$$e^{-\mu} \sum_{i,j} p_i b_{ij} y_j \leq \sum_{i,j} p_i a_{ij} y_j$$

for all  $\mathbf{y}$ . Hence

$$e^{-\mu} \mathbf{pB} \leq \mathbf{pA}. \quad (9.9.1)$$

If we substitute the value of  $y(t)$  in (b), we obtain

$$x_i(t+1) = x_i(t) + \sum_{j=1}^m (b_{ij} - a_{ij}) y_j e^{\lambda t}.$$

The solution of this first-order difference equation is given by

$$\begin{aligned} x_i(t) &= x_i(0) + \frac{\sum_{j=1}^m (b_{ij} - a_{ij}) y_j}{1 - e^\lambda} + \frac{e^{\lambda t}}{e^\lambda - 1} \sum_j (b_{ij} - a_{ij}) y_j \\ &= c_0 + \frac{1}{e^\lambda - 1} e^{\lambda t} \sum_{j=1}^m (b_{ij} - a_{ij}) y_j. \end{aligned} \quad (9.9.2)$$

Using this result in (a), we obtain

$$\begin{aligned} c_0 + \frac{1}{e^\lambda - 1} e^{\lambda t} \sum_j (b_{ij} - a_{ij}) y_j &\geq \sum_j a_{ij} y_j e^{\lambda t}, \\ e^{-\lambda t} (e^\lambda - 1) c_0 + \sum_j b_{ij} y_j &\geq e^\lambda \sum_j a_{ij} y_j. \end{aligned}$$

Since this inequality holds for all values of  $t$ , we permit  $t$  to become infinite and the inequality reduces to

$$\sum b_{ij} y_j \geq e^\lambda \sum a_{ij} y_j, \quad \text{or} \quad \mathbf{By} \geq e^\lambda \mathbf{Ay} \quad (9.9.3)$$

In the following section, by imposing some additional mild positivity restrictions on  $\mathbf{A}$  and  $\mathbf{B}$ , we shall establish the existence of nonnegative vectors  $\mathbf{p}^0, \mathbf{y}^0$  fulfilling (9.9.3) and (9.9.1) when  $\lambda = \mu > 0$ . In this case we say that the economy undergoes a balanced growth such that the expansion rate and interest rate are equal. For other applications of this model the reader is referred to Problems 22 and 23.

**9.10 A general model of balanced growth.** In this section we formulate a general model of production which includes the von Neumann model as a particular case, and prove the existence of a unique rate of uniform expansion and its coincidence with the rate of interest. The dynamic Leontief model (see Section 8.6) may also be considered as a special case of the present model, and these consequences will also be indicated.

Consider an economy in which  $n$  distinct commodities are produced, and in which the technology of production obeys the laws of nonincreasing marginal rates of transformation (i.e., the production possibility set is convex) and of constant returns to scale (i.e., the production possibility set is a cone).

The technological possibility of production is described by the transformation set  $T$  of the economy, where  $T$  is a set of pairs of  $n$  vectors  $\{\mathbf{x}, \mathbf{y}\}$  such that the production of output  $\mathbf{y} = \langle y_1, \dots, y_n \rangle$  is technologically possible from input  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  if and only if  $\{\mathbf{x}, \mathbf{y}\} \in T$ .

The transformation set  $T$  is assumed to possess the following properties:

(T<sub>1</sub>)  $T$  is a closed convex cone in the nonnegative orthant of the  $2n$ -vector space.

(T<sub>2</sub>)  $\{\mathbf{x}, \mathbf{y}\} \in T$  and  $\mathbf{x}' \geq \mathbf{x}$ ,  $0 \leq \mathbf{y}' \leq \mathbf{y}$  imply that  $\{\mathbf{x}', \mathbf{y}'\} \in T$ . This assumption states that the disposal activity is costless. A smaller output is certainly possible with a larger input.

(T<sub>3</sub>)  $\{\mathbf{0}, \mathbf{y}\} \in T$  implies that  $\mathbf{y} = \mathbf{0}$ . That is, it is impossible to produce something from nothing.

(T<sub>4</sub>) For any  $i = 1, \dots, n$  there exists  $\{\mathbf{x}^i, \mathbf{y}^i\} \in T$  such that  $y_i^i > 0$ . That is, every commodity can be produced. (T<sub>4</sub>), in view of (T<sub>1</sub>), is equivalent to the following assertion:

(T<sub>4</sub>) There exists  $\{\mathbf{x}^0, \mathbf{y}^0\} \in T$  such that  $\mathbf{y}^0 \gg \mathbf{0}$ .

For a possible nonzero input-output relation  $\{\mathbf{x}, \mathbf{y}\} \in T$ , the rate of expansion  $\lambda(\mathbf{x}, \mathbf{y})$  will be defined by

$$\lambda(\mathbf{x}, \mathbf{y}) = \max \{\lambda | \mathbf{y} \geq \lambda \mathbf{x}\}. \quad (9.10.1)$$

By (T<sub>1</sub>) and (T<sub>3</sub>)

$$0 \leq \lambda(\mathbf{x}, \mathbf{y}) < \infty. \quad (9.10.2)$$

► **THEOREM 9.10.1.** There exists  $\{\mathbf{x}^*, \mathbf{y}^*\} \in T$  such that

$$\mathbf{y}^* = \lambda^* \mathbf{x}^*, \quad \lambda^* = \lambda(\mathbf{x}^*, \mathbf{y}^*) \quad (\mathbf{x}^* > \mathbf{0}), \quad (9.10.3)$$

and

$$\lambda^* \geq \lambda(\mathbf{x}, \mathbf{y}) \quad (9.10.4)$$

for any  $\{\mathbf{x}, \mathbf{y}\} \in T$  with  $\mathbf{x} > \mathbf{0}$ .

*Proof.* Let

$$\lambda^* = \sup_{\substack{\{\mathbf{x}, \mathbf{y}\} \in T \\ \mathbf{x} > \mathbf{0}}} \lambda(\mathbf{x}, \mathbf{y}). \quad (9.10.5)$$

Then  $0 < \lambda^* < \infty$ . The left-hand inequality is easily demonstrated: by taking  $\{\mathbf{x}^0, \mathbf{y}^0\} \in T$  with  $\mathbf{y}^0 \gg \mathbf{0}$ , we obtain

$$\lambda^* \geq \lambda(\mathbf{x}^0, \mathbf{y}^0) > 0.$$

As for the right-hand inequality, if  $\lambda^*$  is not finite, then there exists a sequence  $\{\mathbf{x}^\nu, \mathbf{y}^\nu\} \in T$  such that  $\mathbf{y}^\nu > \mathbf{0}$ ,  $\mathbf{y}^\nu \geq \lambda^\nu \mathbf{x}^\nu$ , and  $\lim \lambda^\nu = \infty$ . Normalizing  $\mathbf{y}^\nu$  such that

$$\sum_{i=1}^n y_i^\nu = 1$$

and taking a limit point  $\bar{\mathbf{y}}$  of  $(\mathbf{y}^\nu)$ , we are led to the existence of  $\{\mathbf{0}, \bar{\mathbf{y}}\} \in T$  such that  $\bar{\mathbf{y}} > \mathbf{0}$ . This contradicts  $(T_3)$ .

Since  $T$  is a closed cone, there exists a vector pair  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} \in T$  with  $\lambda^* = \lambda(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ , where  $\lambda^*$  is defined by (9.10.5). By  $(T_2)$  we can find  $\{\mathbf{x}^*, \mathbf{y}^*\}$  which satisfies (9.10.3) and (9.10.4). This completes the proof of the theorem.

The vector pair  $\{\mathbf{x}^*, \mathbf{y}^*\}$  is referred to as a *balanced growth* and  $\lambda^*$  as the *rate of balanced growth*.

Associated with the balanced growth we have also an equilibrium price system, as indicated in the following theorem.

► **THEOREM 9.10.2.** There exists a vector  $\mathbf{p}^*$  such that  $\mathbf{p}^* > \mathbf{0}$  and

$$(\mathbf{p}^*, \mathbf{y}) \leq \lambda^*(\mathbf{p}^*, \mathbf{x}) \quad (9.10.6)$$

for all  $\{\mathbf{x}, \mathbf{y}\} \in T$ .

*Proof.* Consider a set  $V$  in  $E^n$  defined as follows:

$$V = \left\{ \mathbf{y} - \lambda^* \mathbf{x} \mid \{\mathbf{x}, \mathbf{y}\} \in T, \quad \sum_{i=1}^n (x_i + y_i) \leq 1 \right\}.$$

Since  $T$  is a closed convex set,  $V$  is also a closed convex set. Moreover, from the very meaning of  $\lambda^*$  we note that  $V$  does not overlap the interior of the positive orthant  $O_n$  of  $E^n$ . Hence there exists a hyperplane through the origin which separates  $O_n$  and  $V$  (Theorem B.1.3). That is, there exists a nonzero vector  $\mathbf{p}^* = \langle p_1^*, \dots, p_n^* \rangle$  such that

$$(\mathbf{p}^*, \mathbf{y} - \lambda^* \mathbf{x}) \leq 0 \quad (9.10.7)$$

for all  $\{\mathbf{x}, \mathbf{y}\} \in T$  (the normalization being irrelevant) and

$$(\mathbf{p}^*, \mathbf{z}) \geq 0 \quad (9.10.8)$$

for all  $\mathbf{z}$  in  $O_n$ . It follows readily from (9.10.8) that  $\mathbf{p}^* > 0$ . Relation (9.10.7) is exactly the statement of the theorem.

*The von Neumann model.* The von Neumann model of an expanding economy may be considered as a special case of the present model. We conform to the notation of the preceding section. Let the number of activities be labeled by  $j$  ( $j = 1, \dots, m$ ). Suppose that the operation of the  $j$ th activity at unit level requires an input  $\langle a_{1j}, \dots, a_{nj} \rangle$  of the  $n$  commodities and produces  $\langle b_{1j}, \dots, b_{nj} \rangle$ . It will be assumed that

$$(i) \quad a_{ij} \geq 0, \quad b_{ij} \geq 0 \quad (i = 1, \dots, n; j = 1, \dots, m).$$

The transformation set  $T$  is now defined by

$$T = [\{\mathbf{x}, \mathbf{y}\} \mid \mathbf{x} \geq \mathbf{Az}, \mathbf{y} \leq \mathbf{Bz} \text{ for some } \mathbf{z} \geq \mathbf{0}, \mathbf{z} = \langle z_1, \dots, z_m \rangle],$$

where

$$\mathbf{A} = \|a_{ij}\|, \quad \mathbf{B} = \|b_{ij}\| \quad (i = 1, \dots, n; j = 1, \dots, m).$$

Assumptions (T<sub>1</sub>) and (T<sub>2</sub>) are always satisfied. (T<sub>3</sub>) is equivalent to:

(ii) For any  $j$  there is at least one  $i$  such that  $a_{ij} > 0$ . (Every activity uses some commodity as input.)

(T<sub>4</sub>) is equivalent to:

(iii) For any  $i$ , there is at least one  $j$  such that  $b_{ij} > 0$ . (Every commodity can be produced by some activity.)

Theorem 9.10.1 combined with Theorem 9.10.2 implies the following theorem.

► **THEOREM 9.10.3.** If  $\mathbf{A}$  and  $\mathbf{B}$  satisfy conditions (i)–(iii), then there exist an activity level and a price system  $\mathbf{p}^*$  such that

$$\mathbf{z}^* > \mathbf{0}, \quad \mathbf{p}^* > \mathbf{0}; \quad \mathbf{Bz}^* \geq \lambda^* \mathbf{Az}^*; \quad \mathbf{p}^* \mathbf{B} \leq \lambda^* \mathbf{p}^* \mathbf{A}.$$

*The Leontief dynamic model.* Suppose that there are  $n$  commodities as before. Let  $\mathbf{A}$  be the input-output coefficient matrix; an element  $a_{kj}$  of  $\mathbf{A}$  represents the amount of commodity  $k$  required for one unit of output of commodity  $j$ . Let  $\mathbf{B}$  denote the capital coefficient matrix. Here  $b_{kj}$  can be interpreted as the amount of commodity  $k$  the producer must have on hand in order to produce one unit of commodity  $j$ .

A commodity vector  $\mathbf{x} = \langle x_1, \dots, x_n \rangle$  at the beginning of the time period will produce a commodity vector  $\mathbf{y} = \langle y_1, \dots, y_n \rangle$  at the end of the period if and only if for some activity level  $\mathbf{z} = \langle z_1, \dots, z_n \rangle$ ,

$$\mathbf{y} \leq (\mathbf{I} - \mathbf{A})\mathbf{z} + \mathbf{x}, \quad \mathbf{x} \geq \mathbf{Bz}, \quad \mathbf{z} > \mathbf{0}. \quad (9.10.9)$$

We have here identified activity level with output. (This merely amounts to an appropriate choice of units of measurement.)

There is an exogenous source of input—for example, raw material or land—such that the stock of commodities grows with each time period. Relations (9.10.9) express in a sense this growth process within the framework of the natural economic limitations of availability and producibility.

Relations (9.10.9) can be elucidated as follows. Stock on hand of the  $k$ th commodity plus output of the  $k$ th commodity in a given time period is given by the formula  $x_k + z_k$ . During the same period an amount  $\sum_j a_{kj} z_j$  of the  $k$ th commodity is used up; hence  $x_k + z_k - \sum_j a_{kj} z_j$  remains available for final consumption. The capital constraints  $\mathbf{x} \geq \mathbf{Bz}$  do not affect this amount, since the investment in equipment remains a capital asset.

In order to deal with the mathematical aspects of this model, it seems convenient to assume that there exists a vector  $\mathbf{z}^0$  ( $\mathbf{z}^0 > \mathbf{0}$ ) such that

$$(\mathbf{I} - \mathbf{A})\mathbf{z}^0 \gg \mathbf{0}. \quad (9.10.10)$$

This assumption asserts the existence of an activity vector which produces a surplus of each commodity; it is vital to the correctness of the relations (9.10.12a,b) below, which in fact need not hold if we do not require (9.10.10). (See p. 257 for a further discussion of the relevance of this kind of assumption.)

We shall further assume that for each  $j$  there exists an  $i$  such that

$$b_{ij} > 0. \quad (9.10.11)$$

The transformation set  $T$  in this case is defined by

$$T = \{(\mathbf{x}, \mathbf{y}) | \mathbf{0} \leq \mathbf{y} \leq (\mathbf{I} - \mathbf{A})\mathbf{z} + \mathbf{x}, \quad \mathbf{x} \geq \mathbf{Bz} \quad \text{for some} \quad \mathbf{z} \geq \mathbf{0}\}.$$

With (9.10.10) and (9.10.11) assumed to hold, it becomes evident that assumptions (T<sub>1</sub>)–(T<sub>4</sub>) are satisfied. It is also easily seen that Theorems 9.10.1 and 9.10.2 imply the existence of  $\lambda^* > 1$ ,  $\mathbf{p}^* > \mathbf{0}$ ,  $\mathbf{z}^* > \mathbf{0}$  such that

$$(\mathbf{I} - \mathbf{A})\mathbf{z}^* \geq (\lambda^* - 1)\mathbf{Bz}^*, \quad (9.10.12a)$$

$$\mathbf{p}^*(\mathbf{I} - \mathbf{A}) \leq (\lambda^* - 1)\mathbf{p}^*\mathbf{B}. \quad (9.10.12b)$$

These last two inequalities can be converted to equalities by imposing further assumptions on the matrix  $\mathbf{A}$ . We indicate one such extension. Two indices  $i$  and  $j$  are said to communicate (in symbols,  $i \sim j$ ; the relation may be nonreflexive) if there exists a chain of indices  $i_1, i_2, \dots, i_r$ , such that  $a_{i_\nu i_{\nu+1}} > 0$  for  $\nu = 1, \dots, r - 1$ , where  $i_1 = i$  and  $i_r = j$ . If all

indices communicate, then the relations (9.10.12) may be written with equality, and  $\mathbf{z}^*$  and  $\mathbf{p}^*$  are strictly positive vectors. To begin with, we prove that  $\mathbf{z}^* \gg \mathbf{0}$ . Let us suppose to the contrary that  $z_{i_0}^* = 0$  for some  $i_0$ . Examining the  $i_0$ th equation of (9.10.12a), we conclude that  $z_i^* = 0$  for all  $i$  which communicate with  $i_0$  for a chain of length 1. Iteration shows that  $z_i^* = 0$  for all  $i$  which communicate with  $i_0$ . This gives  $\mathbf{z}^* \equiv \mathbf{0}$ , contrary to the construction of (9.10.12a); hence  $\mathbf{z}^* \gg \mathbf{0}$ .

It follows that equality holds in (9.10.12b). For otherwise

$$(\mathbf{p}^*(\mathbf{I} - \mathbf{A}), \mathbf{z}^*) < (\lambda^* - 1)(\mathbf{p}^*\mathbf{B}, \mathbf{z}^*). \quad (9.10.13)$$

But premultiplication of (9.10.12a) by  $\mathbf{p}^*$  gives

$$(\mathbf{p}^*(\mathbf{I} - \mathbf{A}), \mathbf{z}^*) \geq (\lambda^* - 1)(\mathbf{p}^*\mathbf{B}, \mathbf{z}^*), \quad (9.10.14)$$

which is obviously incompatible with (9.10.13).

Since equality holds in (9.10.12b) we may argue as above and show that  $\mathbf{p}^* \gg \mathbf{0}$ . By a symmetrical argument this yields

$$(\mathbf{I} - \mathbf{A})\mathbf{z}^* = (\lambda^* - 1)\mathbf{Bz}^*,$$

and the proof is complete. (For further results of this kind, see Problems 20 and 21.)

*Efficient production.* In any model in which the transformation set  $T$  satisfies assumptions (T<sub>1</sub>)–(T<sub>4</sub>) the efficiency of a production plan is easily characterized by the following definition:

A possible production plan  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is said to be efficient if  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\} \in T$ ,  $\bar{\mathbf{x}} > \mathbf{0}$  and there is no  $\{\mathbf{x}, \mathbf{y}\} \in T$  distinct from  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  such that  $\mathbf{x} \leq \bar{\mathbf{x}}$ ,  $\bar{\mathbf{y}} \leq \mathbf{y}$ .

► **THEOREM 9.10.4.** A production plan  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is efficient if and only if there exist price systems  $\bar{\mathbf{p}}$  and  $\bar{\mathbf{q}}$  with positive components  $(\bar{\mathbf{p}}, \bar{\mathbf{q}} \gg \mathbf{0})$  such that

$$(\bar{\mathbf{q}}, \mathbf{y}) - (\bar{\mathbf{p}}, \mathbf{x}) \leq (\bar{\mathbf{q}}, \bar{\mathbf{y}}) - (\bar{\mathbf{p}}, \bar{\mathbf{x}}) \quad (9.10.15)$$

for all  $\{\mathbf{x}, \mathbf{y}\} \in T$ .

*Remark 9.10.1.* The left-hand expression evaluates the net profit—i.e., the value of the output  $(\bar{\mathbf{q}}, \mathbf{y})$  minus the cost of the input  $(\bar{\mathbf{p}}, \mathbf{x})$ —according to the price system  $\{\bar{\mathbf{q}}, \bar{\mathbf{p}}\}$ . The theorem asserts that the principle of profit maximization applies with some appropriate price vector for each efficient production plan.

*Proof of Theorem 9.10.4.* We consider the set  $\Gamma$  in  $E^{2n}$  defined as follows:

$$\Gamma = \langle \lambda \bar{\mathbf{x}} - \mathbf{x}, \quad \mathbf{y} - \lambda \bar{\mathbf{y}} \rangle$$

for  $\{\mathbf{x}, \mathbf{y}\} \in T$  and  $\lambda \geq 0$ . The set  $\Gamma$  is a closed convex cone which touches the nonnegative orthant  $O_{2n}$  only at the origin. Moreover, the set  $\Gamma$  cannot be tangent at the origin to any direction in the positive orthant. Otherwise, since  $\Gamma$  is a closed cone, the ray in that direction of the first orthant would belong to  $\Gamma$ , which is impossible by the definition of  $\Gamma$ . Consequently, we can find a hyperplane through the origin, containing no direction of  $O_{2n}$ , which separates  $\Gamma$  and  $O_{2n}$ . We denote the components of the normal to this hyperplane by  $\langle \bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_n; \bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2, \dots, \bar{\mathbf{q}}_n \rangle$ . Since the hyperplane supports  $O_{2n}$ , it follows easily that  $\bar{p}_i \geq 0$  and  $\bar{q}_i \geq 0$ . Actually,  $\bar{p}_i > 0$  and  $\bar{q}_i > 0$ , since the hyperplane contains no segment of  $O_{2n}$ . Finally,

$$(\bar{\mathbf{p}}, \mathbf{x} - \mathbf{x}) + (\bar{\mathbf{q}}, \mathbf{y} - \bar{\mathbf{y}}) \leq 0$$

for  $\{\mathbf{x}, \mathbf{y}\} \in T$ , since the hyperplane separates  $\Gamma$  from  $O_{2n}$ . This establishes (9.10.15).

On the other hand, if there exist systems  $\bar{\mathbf{p}} \gg 0$  and  $\bar{\mathbf{q}} \gg 0$  satisfying (9.10.15), the efficiency of  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is evident.

### 9.11 Problems.

1. Let  $\mathbf{D}$  be a diagonal matrix, and let  $\mathbf{A}$  be a stable matrix (i.e., all eigenvalues of  $\mathbf{A}$  have negative real parts) such that  $\mathbf{D}\mathbf{A}$  is stable if and only if the diagonal elements of  $\mathbf{D}$  are all positive ( $d_{ii} > 0$ ). Let  $\mathbf{C} = \mathbf{E}\mathbf{A}\mathbf{E}^{-1}$ , where  $\mathbf{E}$  is any nonsingular diagonal matrix. Prove that  $\mathbf{C}$  is stable, and that  $\mathbf{DC}$  is stable if and only if  $d_{ii} > 0$  for all  $i$ .

2. Let  $\mathbf{A}$  be any stable matrix. If  $\mathbf{D}$  is diagonal and  $\mathbf{D}\mathbf{A}$  is stable, prove that at least one characteristic root of  $\mathbf{D}$  is positive.

3. Show that if  $\mathbf{A} = \|a_{ij}\|$  is a matrix such that  $a_{ii} < 0$  and

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad \text{for all } i,$$

then  $\mathbf{A}$  is stable.

4. Consider a matrix  $\mathbf{D} = \|d_{ij}\|$  such that  $\text{sign } d_{ij} = \text{sign } d_{ji}$  ( $i \neq j$ ) and  $\text{sign } (d_{ij}, d_{jk}) = \text{sign } d_{ik}$  ( $i \neq j \neq k$ ) provided  $d_{ij}d_{jk} \geq 0$ . Show that  $\mathbf{D}$  is an  $\bar{M}$ -matrix (see p. 308).

The next two problems apply to the normalized price adjustment process, and the following five problems refer to the nonnormalized price adjustment process. We keep the notation of Section 9.2.

5. Consider the case of  $n = 2$  commodities. At any point  $\mathbf{p}^*$  of stable equilibrium show that

$$\frac{\partial F_0}{\partial p_1}(\mathbf{p}^*) \geq 0, \quad \frac{\partial F_1}{\partial p_0}(\mathbf{p}^*) \geq 0.$$

6. Show that for any number of commodities it is impossible to divide all the commodities (including the numéraire) into two nonvoid sets  $R$  and  $S$  such that  $\partial F_r / \partial p_s < 0$  for all  $r \in R, s \in S$  and all price vectors  $\mathbf{p}$ .

In Problems 7–11 we assume that the excess demand functions  $\mathbf{F}(\mathbf{p})$  are continuously differentiable for  $\mathbf{p} \gg \mathbf{0}$  and homogeneous of degree zero.

7. Suppose that  $\mathbf{F}$  has the property of strict gross substitutability. Prove that if  $\mathbf{F}(\mathbf{p}^0) \gg \mathbf{0}$ , then  $\mathbf{F}[\mathbf{p}(t)] \gg \mathbf{0}$ , where  $\mathbf{p}(t) = \psi(t, \mathbf{p}^0)$ .

8. Let strict gross substitutability prevail for the nonnormalized price adjustment process of (9.2.4). Prove that

$$V(t) = \max_{0 \leq i \leq n} |p_i(t)/p_i^* - 1|$$

is strictly decreasing unless  $p_i(t) = \lambda p_i^*$  for all  $i$  and some positive  $\lambda$ .

9. For the case of two commodities, prove that

$$V(t) = [p_0(t) - p_0^*]^{2k} + [p_1(t) - p_1^*]^{2k}$$

is strictly decreasing in  $t$  for any positive integer  $k$ , provided that  $p_i(t) \neq \lambda p_i^*$  for  $i = 0, 1$  and some positive  $\lambda$ . (Assume strict gross substitutability.)

10. If strict gross substitutability prevails for two commodities, show that

$$\sum_{i,j=0}^1 p_i^* \frac{\partial F_i}{\partial p_j}(\mathbf{p}) F_j(\mathbf{p}) < 0$$

unless  $\mathbf{p} = \lambda \mathbf{p}^*$  ( $\lambda > 0$ ).

11. Under the same conditions as in Problem 10, prove that

$$V(t) = \sum_{i=0}^1 [p_i(t) - p_i^*]^2$$

is a strictly decreasing convex function of  $t$ , where  $\mathbf{p}(t) = \psi(t, \mathbf{p}^0)$  and  $\mathbf{p}^0 \neq \lambda \mathbf{p}^*$ .

12. Let  $\mathbf{F}(\mathbf{p})$  be homogeneous, finite in the interior of the positive orthant  $I$ , but continuous in the extended sense (i.e., infinite values allowed) in the closure of  $I$  with the origin removed. Moreover, assume that  $\partial F_i / \partial p_j$  ( $i \neq j$ ) is strictly positive and continuous over the same domain. Prove that if  $\mathbf{p} > \mathbf{0}$  and  $p_k = 0$ , then  $F_k(\mathbf{p}) = \infty$ .

13. Under the same conditions as in Problem 12, with the added restrictions that  $\mathbf{F}(\mathbf{p})$  is uniformly bounded from below and satisfies the Walras law, prove the existence of a strictly positive equilibrium vector.

14. Let strict gross substitutability prevail for the nonnormalized process (9.2.4), and assume the existence of an equilibrium vector  $\mathbf{p}^* \gg \mathbf{0}$ . Show that Theorem 9.5.4 applies if we take for  $\phi(\mathbf{p})$  any one of the following:

- (a)  $\max_{1 \leq i \leq r} \left\{ \frac{p_i(t)}{p_i^*} \right\},$
- (b)  $\sum_{i=1}^r \max (0, p_i(t) F_i[\mathbf{p}(t)]),$
- (c)  $\max_{1 \leq i \leq r} \left\{ \frac{F_i[\mathbf{p}(t)]}{p_i(t)} \right\}.$

15. Consider the case of the nonnormalized process of two commodities such that the Walras law is satisfied and the excess demand functions are homogeneous of degree zero. Assume that  $F_1(p_1, 1) < 0$  if  $p_1$  is sufficiently large. Define

$$\phi(p_1, p_2) = - \int_0^{p_1/p_2} F_1(v, 1) dv \quad \text{for } p_1, p_2 > 0.$$

Show that Theorem 9.5.4 applies and hence that the process (9.2.5) is *quasi-stable*.

16. Let  $F_i$  denote the excess demand functions and suppose that  $\partial F_i / \partial p_j$  is symmetric at an equilibrium price vector  $\mathbf{p}^* \gg \mathbf{0}$ . Show that local stability of the system is invariant with respect to the choice of numéraire  $i = 0, 1, \dots, r$ .

17. Prove that if the excess demand function  $\mathbf{F}(\mathbf{p})$  of Section 9.6 is continuous, satisfies the weak axiom of revealed preference, and is homogeneous of order zero, the system

$$\mathbf{p}(t+1) = \max \{0, \mathbf{p}(t) + \rho \mathbf{F}[\mathbf{p}(t)]\} \quad (t = 0, 1, 2, \dots),$$

with nonnegative initial price  $\mathbf{p}(0)$ , is stable provided  $\rho$  is a sufficiently small positive number.

18. Let  $\mathbf{F}(\mathbf{p})$  be an excess demand function. Assume that  $\mathbf{F}(\mathbf{p})$  is differentiable, is homogeneous of order zero, satisfies the gross substitutability condition

$$\frac{\partial F_i}{\partial p_j} > 0 \quad (i \neq j, \mathbf{p} \gg \mathbf{0}),$$

and has an equilibrium vector  $\mathbf{p}^* \gg \mathbf{0}$  such that  $\mathbf{F}(\mathbf{p}^*) = \mathbf{0}$ . Consider  $\mathbf{p}(t)$  ( $t = 1, 2, \dots$ ) defined recursively by

$$F_i[p_1(t+1), \dots, p_i(t+1), p_i(t), \dots, p_r(t)] = 0 \\ (t = 0, 1, 2, \dots; i = 1, 2, \dots, r),$$

where  $\mathbf{p}(0) \gg \mathbf{0}$ . Show first that  $\mathbf{p}(t)$  is well defined, and then that  $\mathbf{p}(t)$  converges to an equilibrium vector.

19. Consider a single commodity whose output in each period (time  $t$ ) is split into two parts: consumption for the next period,  $c(t + 1)$ ; and input for the next period's output,  $x(t + 1)$ . Suppose that one unit of input gives rise in the next period to  $a$  units of output and that the law of constant returns to scale prevails. Suppose that at time  $t = 0$  we start with  $K$  units of the commodity. Prove the formula

$$c(n) + x(n) \leq a^n K - \sum_{r=0}^{n-1} a^{n-r} c(r),$$

and set up the problem of maximizing  $c(n)$  as a linear programming problem.

20. Show that (9.10.12a,b) become equalities if there exists a power  $n_0$  such that  $\mathbf{A}^{n_0} \gg \mathbf{0}$ .

21. Show that (9.10.12a,b) become equalities if there exists a power  $n_0$  such that  $\mathbf{B}^{n_0} \gg \mathbf{0}$ .

22. Consider the von Neumann model of an expanding economy as described in Section 9.9. Prove that if  $a_{ij} + b_{ij} > 0$  for all  $i, j = 1, \dots, n$ , then

$$\sup_{\mu} e^{-\mu} = \inf_{\lambda} e^{-\lambda},$$

where  $e^{-\mu}$  satisfies (9.9.1) for some price vector  $\mathbf{p}$  and  $e^{-\lambda}$  satisfies (9.9.3) for some commodity vector  $\mathbf{y}$ .

23. Consider the model of technological production described by a set of pairs of vectors  $\{\mathbf{x}, \mathbf{y}\} \in T$ , with  $\mathbf{x}, \mathbf{y} \in E^n$  satisfying postulates (T<sub>1</sub>)–(T<sub>4</sub>) of Section 9.10. Suppose in addition that  $x_i + y_i > 0$  ( $i = 1, \dots, n$ ) for each  $\{\mathbf{x}, \mathbf{y}\} \in T$ . Let  $\mathcal{P}$  denote the set of all normalized price vectors ( $\mathbf{p} = \langle p_1, \dots, p_n \rangle$ ,  $p_i \geq 0$ ,  $\sum p_i = 1$ ), and define

$$\phi(\mathbf{p}; \mathbf{x}, \mathbf{y}) = \begin{cases} \frac{\sum_{i=1}^n x_i p_i}{\sum_{i=1}^n y_i p_i} & \sum y_i p_i > 0, \\ 0 & \sum y_i p_i = 0 \end{cases}$$

for  $\mathbf{p} \in \mathcal{P}$  and  $\{\mathbf{x}, \mathbf{y}\} \in T$ . Prove that there exist a vector  $\mathbf{p}_0 \in \mathcal{P}$  and a pair  $\{\mathbf{x}^0, \mathbf{y}^0\} \in T$  such that

$$\phi(\mathbf{p}; \mathbf{x}^0, \mathbf{y}^0) \geq \phi(\mathbf{p}^0; \mathbf{x}^0, \mathbf{y}^0) \geq \phi(\mathbf{p}^0; \mathbf{x}, \mathbf{y})$$

for every  $\{\mathbf{x}, \mathbf{y}\} \in T$  and  $\mathbf{p}^0 \in \mathcal{P}$ .

## NOTES AND REFERENCES TO CHAPTER 9

**§ 9.1.** A good review of the nature of the problem of welfare economics can be found in Samuelson [210]. A critical essay discussing the relations of welfare economics and equilibrium theory is included in Koopmans [147]. Arrow [4] is another important contribution to the area of welfare economics.

The fundamental articles of Arrow [4] and Debreu [62] both offer detailed mathematical analyses of the connections between Pareto optima and equilibrium theory. Debreu's presentation is in very abstract form and applies to infinite-dimensional analogs. Arrow's approach is geometric, appealing frequently to the structure theorems of convex sets. We take still another line in this section: that of reducing the relationship between a given Pareto optimum and the existence of an equilibrium to the vector maximum problem of nonlinear programming discussed in Section 7.4.

Both Arrow and Debreu develop the characterization of Pareto optima for the more general case in which the utility indicator functions  $U_i$  are assumed to be quasi-concave. The proofs in this section are valid when the  $U_i$  are quasi-like concave (see [4]).

**§ 9.2.** Stability of the price adjustment process was studied in the nineteenth century by Walras [252]. In modern form, stability of the dynamic process has been discussed by Hicks [121], Samuelson [210], Metzler [180], Allais [1], and others.

In 1958 Arrow and Hurwicz [11] made a comprehensive survey of stability theory and considerably extended the scope of the then existing results. This paper was in the main superseded by Arrow, Block, and Hurwicz [6], in which the results obtained are considerably sharper. Much of our formulation of the process is based on their presentation, and on several elegant refinements and extensions obtained by Uzawa.

The setup of a discrete-time price adjustment process referred to on p. 326 was suggested by Uzawa.

**§ 9.3.** Part (a) of Theorem 9.3.1 is due originally to Arrow and Hurwicz [11], who gave a rather elaborate proof of its validity. Hahn [117] offered a simpler proof. Our proof is still simpler. Part (b) of this same theorem may be new.

Theorem 9.3.2 is new.

The concepts of "gross substitutes" and "complementary commodities" were introduced by Mosak [185] and Metzler [180].

The relations between  $M$ -matrices and  $\tilde{M}$ -matrices were worked out by Morishima [183].

Theorem 9.3.4 is an application to a local situation of a more general theorem formulated by Arrow, Block, and Hurwicz [6].

**§ 9.4.** Theorems 9.4.1 and 9.4.2 are due to Arrow, Block, and Hurwicz [6]. The elegant analysis of Theorems 9.4.3 and 9.4.4 was devised by Uzawa. A weaker version of Theorem 9.4.3 is contained in the first reference.

**§ 9.5.** Theorem 9.5.1 is new; it is the dual of Theorem 9.5.2, which is due to Arrow, Block, and Hurwicz [6]. Theorem 9.5.3 represents a slight extension of a result of Arrow and Hurwicz [11, Theorem 4]. The concept of quasi-stability

originated with Uzawa. This concept makes possible a unified approach to the problem of deriving stability under a variety of assumptions on the structure of the price adjustment process.

**§ 9.6.** The model of this section was proposed by Uzawa [233]. He established the existence of an equilibrium vector, and we have followed his proof. The normalized discrete-time process was shown by Uzawa to be stable whenever the weak axiom of revealed preference is satisfied everywhere. Theorem 9.6.1 establishes the same conclusion for the nonnormalized discrete-time process when strict gross substitutability prevails.

**§ 9.7.** The discussion of this section follows Enthoven and Arrow [75] and Arrow and McManus [13].

**§ 9.8.** The material of this section stems from Arrow and Nerlove [14]. However, the proof of Theorem 9.8.1 is new and more direct.

**§ 9.9.** The model of an expanding economy was constructed by von Neumann [239] as a sweeping formulation of the dynamics of production and consumption. Its solution was closely related to the development of the theory of games. Unfortunately, in von Neumann's original analyses he was obliged to assume that  $a_{ij} + b_{ij} > 0$  for all  $i$  and  $j$ —i.e., that every commodity is either consumed or produced, which on economic grounds is not easily justified. Kemeny, Thompson, and Morgenstern [142] and Gale [94] later substantially weakened the hypothesis and were still able to obtain the quantitative link between the rate of expansion and the rate of interest.

Other generalizations have been offered by Georgescu-Roegen [104] and Nikaido [192]. The last model is nonlinear in form.

Another version of balanced growth was presented by Solow and Samuelson [228]. Neisser gave an interpretation to this model and Suits [230] provided a simple mathematical proof.

Morishima [184] discusses the theory of balanced growth in terms of dynamic Leontief models.

**§ 9.10.** The general balanced-growth model examined here was proposed by Gale [94]. Our discussion is a simplification of his treatment.

**§ 9.11.** Problems 5, 6, and 8 are based on Arrow and Hurwicz [11]. Problems 12 and 13 are due to Arrow, Block, and Hurwicz [6]. Problems 14–18 were suggested by Uzawa.

## SOLUTIONS TO PROBLEMS OF CHAPTERS 5-9

### CHAPTER 5

*Problem 2.* Let  $x, y, z$  denote the quantities of total crude oil used in the three production patterns. The problem reduces to that of maximizing  $32x + 15y + 12z$  subject to  $x, y, z \geq 0$ , and the availability constraints become

$$\begin{aligned} x + 2y + 2z &\leq 10, \\ 2x + y + 2z &\leq 15. \end{aligned}$$

The solution is  $x = 15/2$ ,  $y = 0$ ,  $z = 0$  and the maximum profit is 240.

*Problem 4.*  $x_1 = n, x_2 = x_3 = \dots = x_n = 0$ .

*Problem 8.* Since  $\mathbf{b} \geq \mathbf{0}$ , it follows that  $\mathbf{x} = \mathbf{0}$  is feasible for Problem I. Now use the existence theorem (Theorem 5.4.2).

*Problem 11.* The problem can be stated mathematically as follows:

Maximize  $\sum_i \beta_i x_i$  subject to  $x_i \geq 0$  and  $\sum_i \alpha_i x_i \leq M$ . Let

$$\frac{\beta_{i_0}}{\alpha_{i_0}} = \max \frac{\beta_i}{\alpha_i}.$$

The solution has the form  $x_{i_0} = M/\alpha_{i_0}$  and  $x_j = 0$  for  $j \neq i_0$ .

*Problem 12.* Arrange the  $a$ 's and  $b$ 's in order as follows:  $a_{\mu_1} < \dots < a_{\mu_n}$  and  $b_{\nu_1} > \dots > b_{\nu_n}$ . Then the minimum is

$$\sum_{i=1}^n a_{\mu_i} b_{\nu_i}.$$

(This is known as Tchebycheff's inequality.)

*Problem 16.* Consider the set  $\Gamma = \{(\mathbf{A} - \lambda_0 \mathbf{B})\mathbf{x} | \mathbf{x} \geq \mathbf{0}\}$ . The definition of  $\lambda_0$  implies that  $\Gamma$  does not intersect the interior of the positive orthant. By Lemma B.1.3 there exists  $\mathbf{y}^0 > \mathbf{0}$  such that  $(\mathbf{y}^0, (\mathbf{A} - \lambda_0 \mathbf{B})\mathbf{x}) \leq 0$  for all  $\mathbf{x} \geq \mathbf{0}$  or  $\lambda_0 \mathbf{y}^0 \mathbf{B} \geq \mathbf{y}^0 \mathbf{A}$ . Hence  $\mu_0 \leq \lambda_0$ , as claimed.

*Problem 18.* That  $v(\mathbf{b})$  is a concave function of  $\mathbf{b}$  is easily shown. Next suppose that  $\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}$  is a saddle point of  $\phi(\mathbf{x}, \mathbf{y}) = (\mathbf{c}, \mathbf{x}) + (\mathbf{y}, \mathbf{b} - \mathbf{A}\mathbf{x})$ , and that  $\{\bar{\mathbf{x}} + \delta\mathbf{x}, \bar{\mathbf{y}} + \delta\mathbf{y}\}$  is a saddle point of  $\tilde{\phi}(\mathbf{x}, \mathbf{y}) = (\mathbf{c}, \mathbf{x}) + (\mathbf{y}, \mathbf{b} + \Delta\mathbf{b} - \mathbf{A}\mathbf{x})$ . Then

$$(\mathbf{c}, \bar{\mathbf{x}} + \delta\mathbf{x}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}(\bar{\mathbf{x}} + \delta\mathbf{x})) \leq (\mathbf{c}, \bar{\mathbf{x}}) + (\bar{\mathbf{y}}, \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}),$$

and hence

$$\Delta v = (\mathbf{c}, \delta\mathbf{x}) \leq (\bar{\mathbf{y}}, \mathbf{A}\delta\mathbf{x}).$$

But  $\mathbf{A}\bar{\mathbf{x}} + \mathbf{A}\delta\mathbf{x} \leq \mathbf{b} + \Delta\mathbf{b}$  and  $(\bar{\mathbf{y}}, \mathbf{A}\bar{\mathbf{x}}) = (\bar{\mathbf{y}}, \mathbf{b})$  [equation (5.3.10)] implies  $(\bar{\mathbf{y}}, \mathbf{A}\delta\mathbf{x}) \leq (\bar{\mathbf{y}}, \Delta\mathbf{b})$ . Collecting, we have  $\Delta v \leq (\bar{\mathbf{y}}, \Delta\mathbf{b})$ , from which the desired conclusion follows by an appropriate choice of  $\Delta\mathbf{b}$ .

*Problem 19.* The following sequence of easily proved equivalent statements establishes the result:

(i)  $\mathbf{x} \geq 0, (\mathbf{c}, \mathbf{x}) \geq 0$ ;

(ii)  $x_i \geq 0 (i = 1, \dots, n), \sum_{i \in I} c_i x_i \geq \sum_{k \in K} (-c_k) x_k$ ;

(iii) there exists  $\alpha_{ik} \geq 0$  such that

$$c_i x_i \geq \sum_{k \in K} \alpha_{ik} \quad \text{for } i \in I, \quad -c_k x_k = \sum_{i \in I} \alpha_{ik} \quad \text{for } k \in K$$

(this is proved by induction on the number of elements in  $I$  and  $K$ );

(iv) there exists  $\gamma_i$  such that

$$x_i = \gamma_i + \sum_{k \in K} \frac{\alpha_{ik}}{c_i} \quad \text{for } i \in I, \quad x_k = \sum_{i \in I} \frac{\alpha_{ik}}{-c_k} \quad \text{for } k \in K;$$

$$(v) \mathbf{x} = \sum_{i \in I} \gamma_i \alpha_i + \sum_{\substack{i \in I \\ k \in K}} \alpha_{ik} \nu_{ik} + \sum_{j \in J} x_j \beta_j,$$

with  $\gamma_i, \alpha_{ik}$ , and  $x_j$  nonnegative.

*Problem 20.* By virtue of Theorem 5.3.1 and Equation (5.3.10) we have the following relations:

$$(\mathbf{c}, \mathbf{x}_0 + \delta \mathbf{x}_0) + (\mathbf{y}_0, \mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \delta \mathbf{x}_0)) \leq (\mathbf{c}, \mathbf{x}_0) \quad \text{and} \quad (\mathbf{y}_0, \mathbf{b}) = (\mathbf{y}_0, \mathbf{A} \mathbf{x}_0),$$

$$(\mathbf{c} + \delta \mathbf{c}, \mathbf{x}_0) + (\mathbf{y}_0 + \delta \mathbf{y}_0, \mathbf{b} + \delta \mathbf{b} - (\mathbf{A} + \delta \mathbf{A}) \mathbf{x}_0) \leq (\mathbf{c} + \delta \mathbf{c}, \mathbf{x}_0 + \delta \mathbf{x}_0),$$

$$(\mathbf{y}_0 + \delta \mathbf{y}_0, \mathbf{b} + \delta \mathbf{b}) = (\mathbf{y}_0 + \delta \mathbf{y}_0, (\mathbf{A} + \delta \mathbf{A})(\mathbf{x}_0 + \delta \mathbf{x}_0)).$$

By combining these relations we obtain the first inequality of the problem. The second is established by using the other half of the saddle-value inequalities.

*Problem 21.* The solution follows an argument parallel to that of Lemma 5.3.1.

*Problem 22.* Compare with Problem C'' of Section 5.11.

*Problem 25.* The problem is to minimize

$$c = \sum_{i=1}^n z_i + \lambda X$$

subject to the constraints  $z_i \geq t_i - X$  and  $0 \leq z_i \leq r_i$ . A solution exists only if  $r_i \geq t_i - X$ , a condition that we assume is satisfied. For fixed  $X$ ,  $c$  is minimized when  $z_i = \max \{0, t_i - X\}$  ( $i = 1, \dots, n$ ), and in this case

$$c(X) = \sum_{k=1}^n \max \{0, t_k - X\} + \lambda X.$$

Suppose for definiteness that  $t_1 \geq t_2 \geq \dots \geq t_n$ . If  $t_m > X \geq t_{m+1}$ , then

$$c(X) = \sum_{k=1}^m t_k + (\lambda - m)X,$$

which is minimized when

$$X = \begin{cases} t_m, & \lambda - m < 0, \\ t_{m+1}, & \lambda - m \geq 0. \end{cases}$$

Hence

$$c(X) = \begin{cases} \sum_1^m t_k + (\lambda - m)t_{m+1}, & \lambda - m \geq 0, \\ \sum_1^m t_k + (\lambda - m)t_m, & \lambda - m < 0. \end{cases}$$

Let  $[\lambda] = m_0$ , where  $m_0 \leq \lambda < m_0 + 1$ ; then the optimum  $X = t_{m_0+1}$  and

$$z_i = \begin{cases} t_i - t_{m_0+1}, & i \leq m_0, \\ 0, & i > m_0. \end{cases}$$

## CHAPTER 6

*Problem 2.* Operate Process A at level 0.25 and Process B at level 1.0; do not operate Process C.

*Problem 4.* Use Lemma 6.1.1.

*Problem 6.* Consider  $\mathbf{zA} = \mathbf{0}$ , with the vector  $\mathbf{z}$  represented as

$$\mathbf{z} = \langle x_1, x_2, \dots, x_n, y_1, \dots, y_m \rangle.$$

It follows that  $x_i = -y_j$  for all  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . Hence the null space of  $\mathbf{A}$  is one-dimensional and the rank of  $\mathbf{A}$  is  $n + m - 1$ .

*Problem 8.*

	$0 \leq \lambda < 1$	$1 \leq \lambda < 2$	$2 \leq \lambda$
$x$	10	1	0
$z_1$	0	1	2
$z_2$	0	5	5
$z_3$	0	3	4

*Problem 10.* See Prager [199].

*Problem 11.* Ship 75.

*Problem 13.* Compare with the model of Fig. 6.5 and set  $c_{ij}$  equal to 1 or 0 according to whether  $(i, j)$  is good or bad. If sufficiently large  $a$ 's and  $b$ 's are chosen, their capacity constraints are not relevant.

*Problem 14.* Let  $p$  be a node with arcs  $A_1, \dots, A_s$  directed into the node and arcs  $B_1, \dots, B_s$  directed outwards and limited by a capacity restraint  $c$ . Replace

the node by two nodes labeled  $p$  and  $p'$  such that  $A_1, \dots, A_r$  is directed into  $p$  and  $B_1, \dots, B_s$  is directed out of  $p'$ , and add an arc directed from  $p$  to  $p'$  with capacity constraint  $c$ .

*Problem 16.* The relation  $\mathbf{x}\mathbf{A} = \mathbf{0}$  holds if and only if  $x_i = x_j$  for all  $i, j = 1, \dots, n$ . It follows that the null space of  $\mathbf{A}$  is one-dimensional, and hence the rank of  $\mathbf{A}$  is  $n - 1$ .

*Problem 19.* By Lemma 6.1.1 every extreme point  $\mathbf{x}$  in  $X$  has exactly  $m$  positive components, and conversely. The result now follows from the further observation that  $\sum_i \delta(x_i)$  is a constant over the set of extreme points of  $X$ .

*Problem 20.* Since  $\varphi(\mathbf{x})$  is concave.

*Problem 21.* See Gross [115].

## CHAPTER 7

*Problem 1.* Theorem 7.1.1 with  $X = E^m$ .

*Problem 2.* By virtue of the concavity of  $g$ , we have

$$g(\mathbf{x}) \leq g(\mathbf{x}^0) + (\mathbf{g}_{x^0}, \mathbf{x} - \mathbf{x}^0) \quad (\text{property vii of Section B.4}).$$

Now if  $\mathbf{x}^0$  maximizes  $(\mathbf{g}_{x^0}, \mathbf{x})$  subject to  $f(\mathbf{x}) \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}$ , it follows that  $(\mathbf{g}_{x^0}, \mathbf{x}) \leq (\mathbf{g}_{x^0}, \mathbf{x}^0)$  for  $\mathbf{x}$  satisfying the constraints, and consequently  $g(\mathbf{x}) \leq g(\mathbf{x}^0)$ . On the other hand, if  $g(\mathbf{x}^0) \geq g(\mathbf{x})$  for all  $\mathbf{x}$  satisfying  $\mathbf{x} \geq \mathbf{0}$  and  $f(\mathbf{x}) \geq \mathbf{0}$ , then  $\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0)$  ( $0 \leq t \leq 1$ ) also satisfies the constraints. The maximum of  $g[\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0)] = I(t)$  where  $0 \leq t \leq 1$  is achieved at  $t = 0$ . Since  $I(t)$  is concave, this can occur if and only if  $I'(0) \leq 0$  or  $(\mathbf{g}_{x^0}, \mathbf{x} - \mathbf{x}^0) \leq 0$ , as was to be shown.

*Problem 3.* By Theorem 7.1.1,

$$(\mathbf{p}, \mathbf{x}) + (\mathbf{u}^0, \mathbf{v} - \mathbf{F}(\mathbf{x})) \leq (\mathbf{p}, \mathbf{x}^0) + (\mathbf{u}^0, \mathbf{v} - \mathbf{F}(\mathbf{x}^0)) \leq (\mathbf{p}, \mathbf{x}^0) + (\mathbf{u}, \mathbf{v} - \mathbf{F}(\mathbf{x}^0))$$

for any  $\mathbf{x}, \mathbf{u} \geq \mathbf{0}$ . Hence  $(\mathbf{u}^0, \mathbf{v} - \mathbf{F}(\mathbf{x}^0)) = \mathbf{0}$ , and thus  $(\mathbf{p}, \mathbf{x}) - (\mathbf{u}^0, \mathbf{F}(\mathbf{x})) \leq (\mathbf{p}, \mathbf{x}^0) - (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0))$ . Note that  $\mathbf{F}(\mathbf{x})$  is homogeneous; hence for  $\mathbf{x} = \lambda \mathbf{x}^0$ , with  $\lambda$  an arbitrary positive scalar, we have  $(\mathbf{p}, \mathbf{x}^0) = (\mathbf{u}^0, \mathbf{F}(\mathbf{x}^0)) = (\mathbf{u}^0, \mathbf{v})$ .

*Problem 4.* Theorem 7.1.1 shows that  $\mathbf{x}^*$  is optimum for the quadratic problem if and only if for some  $\mathbf{u}^* \geq \mathbf{0}$

$$\begin{aligned} \mathbf{c} - \epsilon \mathbf{x}^* - \mathbf{u}^* \mathbf{A} &\leq \mathbf{0}, \\ (\mathbf{c} - \epsilon \mathbf{x}^* - \mathbf{u}^* \mathbf{A}, \mathbf{x}^*) &= \mathbf{0}, \\ \mathbf{A} \mathbf{x}^* &\leq \mathbf{b}, \quad \mathbf{x}^* \geq \mathbf{0}, \\ (\mathbf{u}^*, \mathbf{b} - \mathbf{A} \mathbf{x}^*) &= \mathbf{0}. \end{aligned}$$

It follows from these relations that  $\mathbf{x}^*$  is an optimum for the problem: Maximize  $(\mathbf{c} - \epsilon \mathbf{x}^*, \mathbf{x})$  subject to  $\mathbf{A} \mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ . If  $\epsilon$  is sufficiently small,  $\mathbf{x}^*$  is an optimum for the original linear programming problem (by Theorem 5.3.1).

*Problem 5.* For fixed  $n$  the optimum  $\mathbf{x}^n$  is given by  $x_i^{(n)} = T/n$  and

$$u(n) = \sum_{i=1}^n f(x_i^{(n)}) = nf\left(\frac{T}{n}\right).$$

The maximum of  $u(n)$  is achieved where  $n$  is determined by the condition  $u'(n) = 0$ , or equivalently by

$$\frac{T}{n}f'\left(\frac{T}{n}\right) = f\left(\frac{T}{n}\right).$$

*Problem 6.* Since  $f$  is concave, the optimum is located at an extreme point. Hence  $\min_i \sum f(x_i) = f(T)$ .

*Problem 7.* Use the solution of Problem 5, making sure that the constraints are satisfied.

*Problem 9.* By Theorem 7.1.1,  $\mathbf{x}^*$  is optimum for the first problem if and only if for some  $\mathbf{u}^* \geq 0$

$$\begin{aligned} \mathbf{a} - \mathbf{Ax}^* &= \mathbf{u}^* \mathbf{B}, \\ \mathbf{Bx}^* &\leq \mathbf{b}, \\ (\mathbf{u}^*, \mathbf{b} - \mathbf{Bx}^*) &= 0. \end{aligned}$$

These conditions are equivalent to

$$\begin{aligned} \mathbf{x}^* &= \mathbf{A}^{-1}\mathbf{a} - \mathbf{A}^{-1}\mathbf{B}'\mathbf{u}^* \quad (\mathbf{B}' \text{ is the transpose of } \mathbf{B}), \\ \mathbf{B}\mathbf{A}^{-1}\mathbf{a} - \mathbf{b} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}'\mathbf{u}^* &\leq 0, \\ (\mathbf{B}\mathbf{A}^{-1}\mathbf{a} - \mathbf{b} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}'\mathbf{u}^*, \mathbf{u}^*) &= 0. \end{aligned}$$

These last conditions are equivalent to saying that  $\mathbf{u}^*$  maximizes

$$(\mathbf{B}\mathbf{A}^{-1}\mathbf{a} - \mathbf{b}, \mathbf{u}) - \frac{1}{2}(\mathbf{u}, \mathbf{B}\mathbf{A}^{-1}\mathbf{B}'\mathbf{u})$$

subject to  $\mathbf{u} \geq 0$ , again by Theorem 7.1.1.

*Problem 10.* See Problem 17 of Chapter 5.

*Problem 11.* Use Theorem 7.4.1.

*Problem 12.* Suppose the contrary; then  $\mathbf{F}(\mathbf{x})$  has some component nonpositive for each  $\mathbf{x}$  in  $X$ . Let  $H$  be the image in  $E^n$  of the vector function  $\mathbf{F}(\mathbf{x})$  as  $\mathbf{x}$  ranges over  $X$ . Since  $\mathbf{F}$  is concave, the convex hull of  $H$  does not intersect the interior of the first orthant. Hence a separating plane  $\mathbf{u}$  exists and  $(\mathbf{u}, \mathbf{F}(\mathbf{x})) \leq 0$  for all  $\mathbf{x}$  in  $X$ , contradicting the hypothesis.

*Problem 13.* Let  $g(\mathbf{x}) = (\mathbf{c}, \mathbf{x}) - \frac{1}{2}(\mathbf{x}, \mathbf{Ax})$ . Then  $x_i^{v+1}$  uniquely maximizes  $g(x_1^{(v+1)}, \dots, x_{i-1}^{(v+1)}, x_i, x_{i+1}^{(v)}, \dots, x_n^{(v)})$  subject to  $x_i \geq 0$ . Therefore

$$\begin{aligned} g(x_1^{(v+1)}, \dots, x_{i-1}^{(v+1)}, x_i^{(v+1)}, x_{i+1}^{(v)}, \dots, x_n^{(v)}) \\ \geq g(x_1^{(v+1)}, \dots, x_{i-1}^{(v+1)}, x_i^{(v)}, x_{i+1}^{(v)}, \dots, x_n^{(v)}), \end{aligned}$$

with equality only if  $x_i^{v+1} = x_i^v$ . Thus  $g(x_1^{(v+1)}, \dots, x_n^{(v+1)}) \geq g(x_1^{(v)}, \dots, x_n^{(v)})$ , with strict inequality unless  $x_1^{(v)}, \dots, x_n^{(v)}$  is the maximum of  $g(\mathbf{x})$ . But the set  $\{\mathbf{x} | g(\mathbf{x}) \geq g(\mathbf{x}^{(1)})\}$  is bounded, and the sequence  $g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(2)}), \dots$  is also bounded. Hence  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$  converges to the unique optimum.

*Problem 16.* There exists a vector  $\mathbf{x}^0$  such that  $c = \sup_\alpha f_\alpha(\mathbf{x}^0) < \infty$  if and only if  $\{c, \mathbf{x}^0\} \in [f_\alpha, C_\alpha]$  for all  $\alpha$ , which in turn is equivalent to

$$\sup_{\xi \in \Gamma_\alpha} [(\mathbf{x}^0, \xi) - \varphi_\alpha(\xi)] = f_\alpha(\mathbf{x}^0) \leq c.$$

*Problem 18.* Set  $f(\mathbf{x}) = \min_y (\mathbf{x}, \mathbf{Ay})$  and  $g(\mathbf{y}) = \max_x (\mathbf{x}, \mathbf{Ay})$ , where  $\mathbf{x}$  and  $\mathbf{y}$  traverse the usual sets of simplicities. Apply Theorem 7.8.1.

*Problem 20.* Let  $\mathbf{t}_\alpha$  denote the set of unit vectors through the origin orthogonal to the bounding hyperplanes of  $O_\alpha$  and directed into  $O_\alpha$ . Let  $\mathbf{p}_\alpha$  denote the terminal points of  $\mathbf{t}_\alpha$  on the unit sphere. The hypothesis implies that the convex span of  $\mathbf{p}_\alpha$  has the origin on its interior. Apply Theorem 7.9.4.

*Problem 22.* Consider the set of open half spaces complementary to the specified closed half-spaces  $M_\alpha$ , and apply Problem 21.

*Problem 23.* Identify the rays  $\mathbf{r}_\alpha$  with the points  $\mathbf{p}_\alpha$  at which they pierce the unit sphere centered on the origin. The hypothesis implies that  $\mathbf{p}_\alpha$  spans a convex set with the origin as an interior point. Apply Theorem 7.9.4.

*Problem 24.* A vector  $\mathbf{p} \in S$  will be denoted by  $\langle x_1, \dots, x_n \rangle$ , a vector  $\mathbf{p}' \in S'$  by  $\langle y_1, \dots, y_n \rangle$ . Vectors in  $E^{n+1}$  will be denoted by  $\mathbf{a} = \langle a_0, a_1, \dots, a_n \rangle$ . Let

$$\begin{aligned} H_p &= \{\mathbf{a} | a_0 + a_1 x_1 + \dots + a_n x_n < 0\}, \\ H_{p'} &= \{\mathbf{a} | a_0 + a_1 y_1 + \dots + a_n y_n > 0\}. \end{aligned}$$

We have thus generated a finite subfamily of half-spaces in  $E^{n+1}$ ; any  $n+2$  of them correspond to  $n+2$  points from  $S \cup S'$ . By assumption they intersect. Apply Corollary 7.9.1.

*Problem 25.* Use the method of Problem 24 and Theorem 7.9.4.

## CHAPTER 8

$$\text{Problem 1. } p(t) = \left( p(0) + \frac{\alpha - \beta}{a - b} \right) \exp [-(a - b)t/a] - \frac{\alpha - \beta}{a - b}.$$

If  $a > b$ , then  $p(t)$  is convergent; if  $a = b$ , then  $p(t)$  is a constant; and if  $a < b$ , then  $p(t)$  is divergent.

*Problem 4.*  $Y(t) = cY(t) + v\dot{Y}(t) + A_0 e^{rt}$ . Then

$$Y(t) = \frac{A_0/v}{\frac{1-c}{v} - r} e^{rt} + k e^{(1-c)t/v}.$$

*Problem 6.*  $Y(t) = c_1 Y(t-1) + c_2 Y(t-2) + v(Y_{t-1} - Y_{t-2}) + A$ . Let  $\bar{Y} = A/(1 - c_1 - c_2)$ . Then  $Y(t) = \bar{Y} + k_1 \lambda_1^t + k_2 \lambda_2^t$ , where  $\lambda_1$  and  $\lambda_2$  are roots of  $x^2 - (c_1 + v)x + (v - c_2) = 0$ . If multiple roots or complex roots occur, the form of the solution is modified accordingly.

*Problem 8.* (Proof by induction.) Consider the equation  $\mathbf{D}(\mathbf{I} - \mathbf{C}) = \mathbf{E}$  for  $c_{11} \neq 1$ , where  $d_{11} = 1$ ,  $d_{j1} = c_{j1}/(1 - c_{11})$  for  $j = 2, \dots, n$ , and  $d_{ij} = \delta_{ij}$  otherwise. Then

$$\mathbf{E} = \begin{vmatrix} 1 - c_{11} & -c_{12} & \cdots & -c_{1n} \\ 0 & 1 - \tilde{c}_{22} & \cdots & -\tilde{c}_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & -\tilde{c}_{n2} & \cdots & 1 - \tilde{c}_{nn} \end{vmatrix}$$

where

$$\tilde{c}_{ij} = c_{ij} + \frac{c_{i1}c_{1j}}{1 - c_{11}} \quad \text{for } i, j = 2, \dots, n.$$

Observe that

$$\sum_{i=2}^n |\tilde{c}_{ij}| \leq 1.$$

Hence it follows from the induction hypothesis that

$$\begin{aligned} \det(\mathbf{I} - \mathbf{C}) &= (1 - c_{11}) \det(\mathbf{I} - \tilde{\mathbf{C}}) \geq (1 - |c_{11}|) \prod_{j=2}^n \left(1 - \sum_{i=2}^n |\tilde{c}_{ij}|\right) \\ &\geq \left(1 - \sum_{i=1}^n |c_{i1}|\right) \prod_{j=2}^n \left(1 - \sum_{i=1}^n |c_{ij}|\right). \end{aligned}$$

The modification of this argument where  $c_{11} = 1$  is trivial.

*Problem 10.* Let  $\mathbf{x}^0 \gg \mathbf{0}$  be defined by  $\mathbf{Ax}^0 = \lambda_0(\mathbf{A})\mathbf{x}^0$  and  $\mathbf{Ay} = \mu\mathbf{y}$ , with  $\mathbf{y}$  a real vector and  $\mu$  real. If the  $k$ th coordinate of  $\mathbf{y}$  is zero, then  $\mathbf{A}_{kk}|\mathbf{y}| \geq |\mu| |\mathbf{y}|$ , where  $\mathbf{A}_{kk}$  is the matrix obtained by deleting the  $k$ th row and the  $k$ th column. By Corollary 8.2.1, we have  $\lambda_0(\mathbf{A}) \geq \lambda_0(\mathbf{A}_{kk}) \geq |\mu| \geq \mu$ . If the  $k$ th coordinate of  $\mathbf{y}$  is not zero, we determine  $a$  such that  $a y_k + x_k^0 = 0$ . Since  $\mathbf{y}$  and  $\mathbf{x}$  are linearly independent,  $a\mathbf{y} + \mathbf{x}^0 \neq \mathbf{0}$ . Now  $\mathbf{A}_{kk}(a\mathbf{y} + \mathbf{x}^0) = \mathbf{A}(a\mathbf{y} + \mathbf{x}^0) \geq \mu(a\mathbf{y} + \mathbf{x}^0)$ , and the conclusion follows as before, since  $k$  is arbitrary.

*Problem 11.* See Gantmacher [101, chap. 13].

*Problem 13.* Suppose for definiteness that the first column vector  $\mathbf{b}$  of  $\mathbf{B}(\lambda)$  is nonnegative and not all zeros. Since  $(\lambda\mathbf{I} - \mathbf{A})(\lambda\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I}$ , we have  $(\lambda\mathbf{I} - \mathbf{A})\mathbf{b} > \mathbf{0}$ . By Theorem 8.2.1 there exists a vector  $\mathbf{y}^0 \gg \mathbf{0}$  such that  $\mathbf{y}^0\mathbf{A} = \lambda_0(\mathbf{A})\mathbf{y}^0$ ; hence  $\lambda(\mathbf{y}^0, \mathbf{b}) > \lambda_0(\mathbf{A})(\mathbf{y}^0, \mathbf{b})$  and  $\lambda > \lambda_0(\mathbf{A})$ , as was to be shown.

*Problem 15.* Let  $\lambda$  be the eigenvalue with largest real part. Then  $\mathbf{x}\mathbf{B} = \lambda\mathbf{x}$  for some  $\mathbf{x} > \mathbf{0}$ . Hence  $0 < (\mathbf{x}, \mathbf{B}\mathbf{z}) = \lambda(\mathbf{x}, \mathbf{z})$ . Since  $(\mathbf{x}, \mathbf{z}) \leq 0$ , we have  $\lambda < 0$ .

*Problem 17.* Assume the contrary. Then the set  $\Gamma$  in  $E^{2n}$  where

$$\Gamma = \left\{ \begin{pmatrix} \mathbf{I}, & -\mathbf{C} \\ \mathbf{C}', & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \middle| \begin{pmatrix} \mathbf{C}' & \mathbf{0} \\ \mathbf{0} & -\mathbf{C} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \geq \mathbf{0} \right\},$$

has no point in common with the interior of the positive cone. The remainder of the argument is standard.

*Problem 18.* Let  $\{\mathbf{p}^*, \mathbf{x}_i^*, \mathbf{y}_j^*\}$  be an equilibrium. Then

$$(\mathbf{p}^*, \mathbf{y}_j^*) = \max_{\mathbf{y}_j \in Y_j} (\mathbf{p}^*, \mathbf{y}_j) \quad (j = 1, \dots, n).$$

Since  $Y_j$  is a cone, we deduce that  $(\mathbf{p}^*, \mathbf{y}_j^*) = 0$  for  $j = 1, \dots, n$ , as claimed.

*Problem 19.* (a) This is easily proved by a geometric argument.

(b) Let  $\mathbf{p}^1 = \lambda \mathbf{p}^0$ ,  $M^1 = \lambda M^0$ ,  $\lambda > 0$ , and  $\mathbf{x}^0 = \mathbf{f}(\mathbf{p}^0, M^0)$ ,  $\mathbf{x}^1 = \mathbf{f}(\mathbf{p}^1, M^1)$ . Then  $(\mathbf{p}^1, \mathbf{x}^0) = \lambda(\mathbf{p}^0, \mathbf{x}^0) = \lambda M^0 = M^1 = (\mathbf{p}^1, \mathbf{x}^1)$ . If  $\mathbf{x}^0 \neq \mathbf{x}^1$ , then by the weak axiom of revealed preference  $(\mathbf{p}^0, \mathbf{x}^0) < (\mathbf{p}^1, \mathbf{x}^1)$ . But  $(\mathbf{p}^0, \mathbf{x}^0) = M^0$  and

$$(\mathbf{p}^0, \mathbf{x}^1) = \frac{1}{\lambda} (\mathbf{p}^1, \mathbf{x}^1) = \frac{1}{\lambda} M^1 = M^0,$$

which is a contradiction.

## CHAPTER 9

*Problem 1.* **DC** stable  $\Leftrightarrow \mathbf{E}^{-1}\mathbf{DCE} = \mathbf{E}^{-1}\mathbf{DEA}$  stable (**D** and **E** commute since both are diagonal matrices)  $\Leftrightarrow \mathbf{E}^{-1}\mathbf{DE}$  has all positive diagonal elements  $\Leftrightarrow d_{ii} > 0$ .

*Problem 4.* Classify  $\{1, \dots, n\}$  according to the equivalence relation

$$i \sim j \Leftrightarrow \text{sign } d_{ij} \geq 0.$$

Then  $i \sim j$  implies  $j \sim i$  and  $i \sim j, j \sim k$  imply  $i \sim k$ .

*Problem 7.* Let

$$\lambda(t) = \min_{1 \leq i \leq r} \frac{p_i(t)}{p_i^*},$$

where  $\mathbf{p}^*$  is a strict positive equilibrium vector. Then

$$\lim_{h \rightarrow 0+} \frac{\lambda(t+h) - \lambda(t)}{h} \geq 0,$$

and hence  $\lambda(t) \geq \lambda(0)$ . See also the proof of Theorem 9.4.3.

*Problem 8.*

$$\overline{\lim}_{h \rightarrow 0+} \frac{V(t+h) - V(t)}{h} \leq \begin{cases} \frac{1}{p_i^*} \frac{dp_i^*}{dt} \text{ for } V(t) = -1 + p_i(t)/p_i^* \\ -\frac{1}{p_i^*} \frac{dp_i^*}{dt} \text{ for } V(t) = 1 - (p_i(t)/p_i^*) \end{cases} = \begin{cases} F_i(t)/p_i^*, \\ -F_i(t)/p_i^*. \end{cases}$$

for an appropriate index  $i$ .

If  $p_i(t) > p_i^*$ , it follows from assertion (I) of the proof of Theorem 9.4.3 that  $F_i(t) \leq 0$ , with equality only at an equilibrium point since the sequence  $p_j(t)/p_j^*$  ( $j = 0, 1, \dots, n$ ) is maximized for the index  $j = i$ . Similarly, if  $i$  is such that  $V(t) = -(p_i(t) - p_i^*)$ , then  $i$  is the index for which  $p_j(t)/p_j^*$  ( $0 \leq j \leq n$ ) is minimized, and therefore  $F_i(t) \geq 0$ , with equality only if  $\mathbf{p}(t)$  is an equilibrium vector. In either case, we obtain

$$\overline{\lim}_{h \rightarrow 0+} \frac{V(t+h) - V(t)}{h} \leq 0,$$

and the remainder of the proof proceeds as in Theorem 9.4.3.

*Problem 12.* Consider  $\mathbf{p}^0 > 0$  of the form  $\mathbf{p}^0 = \langle 0, p_2^0, p_3^0, \dots, p_r^0 \rangle$ . We show that  $F_1(\mathbf{p}^0) = \infty$ . Let  $\mathbf{p}_\lambda = \langle \lambda, p_2^0, p_3^0, \dots, p_r^0 \rangle$ , with  $\lambda > 0$ . From Euler's identity (homogeneity) we obtain

$$\lambda \frac{\partial F_1}{\partial p_1}(\mathbf{p}_\lambda) = - \sum_{k=2}^r p_k^0 \frac{\partial F_1}{\partial p_k}(\mathbf{p}_\lambda) \leq -c < 0,$$

or

$$\lambda \frac{dF_1}{d\lambda} \leq -c < 0 \quad \text{for all } 0 \leq \lambda \leq \lambda_0.$$

Integrating this relation from  $\epsilon$  to  $\lambda_0$  gives

$$F_1(\mathbf{p}_\epsilon) > F_1(\mathbf{p}_{\lambda_0}) - c \log \epsilon + c \log \lambda_0,$$

and as  $\epsilon \rightarrow 0$  the conclusion follows.

*Problem 13.* We restrict attention to the set  $\Delta$  of all vectors  $\mathbf{p}$  satisfying  $p_i \geq 0$  and

$$\sum_{i=1}^r p_i = 1.$$

Consider the functions  $g_k(\mathbf{p}) = e^{-F_k(\mathbf{p})}$  for  $\mathbf{p}$  in  $\Delta$  ( $1 \leq k \leq r$ ). By Problem 12,  $g_k(\mathbf{p})$  are continuous on  $\Delta$  and  $g_k(\mathbf{p}) = 0$  whenever  $p_k = 0$  (and conversely). Let

$$h_k(\mathbf{p}) = \max_{1 \leq j \leq r} \{g_j(\mathbf{p})\} - g_k(\mathbf{p}).$$

Suppose there exists no  $\mathbf{p}$  in  $\Delta$  for which  $h_k(\mathbf{p}) = 0$  ( $k = 1, \dots, r$ ). In that

case, consider the mapping of  $\Delta$  into the boundary of  $\Delta$ , defined as follows:

$$\langle p_k \rangle \rightarrow \left\langle \frac{h_k(\mathbf{p})}{\sum_{i=1}^r h_i(\mathbf{p})} \right\rangle.$$

By virtue of the Brouwer fixed-point theorem (Section C.2) we secure a fixed-point vector  $\mathbf{p}^*$  having at least one zero component. If  $h_{k_0}(\mathbf{p}^*) = p_{k_0}^* = 0$ , then  $g_{k_0}(\mathbf{p}^*) = 0$  and

$$\max_{1 \leq j \leq r} g_j(\mathbf{p}^*) = 0,$$

which means that  $\mathbf{p}^* = 0$ , an absurdity. Consequently, there exists a choice  $\mathbf{p}^*$  with the property  $g_1(\mathbf{p}^*) = g_2(\mathbf{p}^*) = \dots = g_r(\mathbf{p}^*) \neq 0$ . Hence  $F_1(\mathbf{p}^*) = F_2(\mathbf{p}^*) = \dots = F_r(\mathbf{p}^*) \neq \infty$ , which by virtue of the Walras law reduces to  $F(\mathbf{p}^*) = 0$  and  $\mathbf{p}^* \gg 0$ .

*Problem 16.* We may assume without loss of generality that  $\mathbf{p}^* = \langle 1, \dots, 1 \rangle$ . Let

$$a_{ij} = \left( \frac{\partial F_i}{\partial p_j} \right)_{\mathbf{p}^*} \quad (i, j = 0, 1, \dots, r).$$

Assume that the system

$$\frac{dp_i}{dt} = \sum_{j=1}^r a_{ij} p_j \quad (i = 1, \dots, r)$$

is stable (the zeroth component is the numéraire). Since the matrix  $\|a_{ij}\|_1^r$  is symmetric it is negative definite. Consider another commodity, say  $r$ , as the numéraire. By homogeneity,

$$a_{i0} = - \sum_{j=1}^r a_{ij} \quad (i = 1, \dots, r)$$

and since the numéraire is constant

$$a_{00} = - \sum_{i=1}^r a_{i0} = \sum_{i,j=1}^r a_{ij}.$$

Now, for any  $\mathbf{x} = \langle x_0, \dots, x_{r-1} \rangle \neq 0$  we obtain by some simple calculations that

$$\sum_{i,j=0}^{r-1} a_{ij} x_i x_j = \sum_{i,j=1}^r a_{ij} x'_i x'_j,$$

where  $x'_i = x_0 - x_i$  for  $i = 1, \dots, r-1$  and  $x'_r = x_0$ . Hence the matrix  $\|a_{ij}\|_0^{r-1}$  is stable.

*Problem 18.* Let

$$\lambda(t) = \min_{1 \leq i \leq r} \left\{ \frac{p_i(t)}{p_i^*} \right\}, \quad \Lambda(t) = \max_{1 \leq i \leq r} \left\{ \frac{p_i(t)}{p_i^*} \right\}.$$

Then  $-\lambda(t)$  and  $\Lambda(t)$  are nonincreasing functions (strictly if  $\mathbf{p}(t)$  is not an equilibrium vector). From here on, an argument analogous to the proof of Theorem 9.4.3 is applicable.

*Problem 19.* The relation of output and consumption over two successive periods obeys the law  $x(t) + c(t) \leq ax(t-1)$ . By iteration,

$$\begin{aligned} a^{n-t}x(t) + a^{n-t}c(t) &\leq a^{n-(t-1)}x(t-1) \\ \vdots \\ a^n x(0) + a^n c(0) &\leq a^n K. \end{aligned}$$

Hence

$$c(n) + x(n) \leq a^n K - \sum_{t=0}^{n-1} a^{n-t} c(t).$$

Stated as a linear programming problem: Maximize  $c(n)$  subject to the constraints

$$a^t K - \sum_{v=0}^t a^{n-v} c(v) \geq v \quad (t = 0, 1, 2, \dots, n)$$

and  $c(v) \geq 0$  for all  $v$ .

*Problem 20.* In this case  $\mathbf{z}^* \gg \mathbf{0}$  and  $\mathbf{p}^* \gg \mathbf{0}$ .

## APPENDIXES

In order to make this volume as complete in itself as possible and accessible to the mature student who has only a modicum of mathematical training beyond the calculus, we devote these appendixes to a brief review of those topics of upper-class undergraduate mathematics which play a fundamental role in the studies of this book. Where an analysis requires more advanced mathematics, the essentials are usually directly incorporated in the text.

In Appendix A we summarize the salient features of the theory of vector spaces and matrices. Our approach to this theory, unlike the usual approach, is primarily geometric rather than algebraic. We do not develop matrix theory abstractly in terms of general rings and fields, but restrict the scalar field of our vector spaces to the real or complex numbers. Proofs that are not easily accessible in the literature are indicated in detail, and in all circumstances we have at least sketched the arguments verifying the main statements of the theory.

Appendix B treats some aspects of the theory of convex sets and convex functions. Much of the theory of finite games, linear programming, and mathematical economics is based on operations with sets and functions derived from linear and convex inequalities. We summarize here the appropriate versions of the separation theorems for convex sets and discuss simple forms of the duality theory of convex cones. Some advanced topics in the theory of convex sets—for example, the development of the theory of conjugate convex functions and certain convexity results connected with moment spaces—are presented in the text.

Appendix C reviews several miscellaneous facts regarding functions, integrals, and fixed-point theorems which are explicitly called upon in making some of the analyses in this volume. The concepts of semi-continuity and equi-continuity are defined precisely, and some of the properties of these functions are listed. The important Kakutani fixed-point theorem is noted in Section C.2. The relevant properties of Lebesgue-Stieltjes integrals are recorded for reference in Section C.3.

The following books and articles may profitably be consulted in connection with the topics covered in these appendixes:

- Appendix A: Perlis [195], Birkoff and McLane [26], Gantmacher [101].
- Appendix B: Fenchel [81], Hardy, Littlewood, and Pólya [118, Chap. 2].
- Appendix C: Graves [112], Rudin [209].

## APPENDIX A

## VECTOR SPACES AND MATRICES

**A.1 Euclidean and unitary spaces.** A real vector space may be defined as follows: We have a set  $V$ ; an associative and commutative addition,  $+$ , of pairs of its elements; a unique zero element  $\theta$ ; and a multiplication of elements of  $V$  by real numbers, for which, with  $\alpha, \beta$  real and  $\mathbf{x}, \mathbf{y}$  in  $V$

$$\begin{aligned} \mathbf{x} + \theta &= \mathbf{x} & 1 \cdot \mathbf{x} &= \mathbf{x}, & 0 \cdot \mathbf{x} &= \theta, \\ \alpha(\mathbf{x} + \mathbf{y}) &= \alpha\mathbf{x} + \alpha\mathbf{y}, & (\alpha + \beta)\mathbf{x} &= \alpha\mathbf{x} + \beta\mathbf{x}, & \alpha(\beta\mathbf{x}) &= (\alpha\beta)\mathbf{x}. \end{aligned}$$

As a prime example we have the Euclidean  $n$ -dimensional space  $E^n$ , which consists of the set of all ordered  $n$ -tuples  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  of real numbers  $x_1, x_2, \dots, x_n$ , with addition and multiplication by a real number  $\gamma$  defined by

$$\begin{aligned} \langle x_1, x_2, \dots, x_n \rangle + \langle y_1, y_2, \dots, y_n \rangle &= \langle x_1 + y_1, x_2 + y_2, \dots, x_n + y_n \rangle, \\ \gamma \langle x_1, x_2, \dots, x_n \rangle &= \langle \gamma x_1, \gamma x_2, \dots, \gamma x_n \rangle. \end{aligned}$$

Here  $\theta$  (which we shall henceforth call  $\mathbf{0}$ ) is  $\langle 0, 0, \dots, 0 \rangle$ ; each  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  has a negative  $-\mathbf{x} = \langle -x_1, -x_2, \dots, -x_n \rangle$  for which  $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$  (the same is true abstractly:  $\mathbf{x} + (-1) \cdot \mathbf{x} = 1 \cdot \mathbf{x} + (-1) \cdot \mathbf{x} = [1 + (-1)] \cdot \mathbf{x} = 0 \cdot \mathbf{x} = \theta$ ); and we may define subtraction by  $\mathbf{y} - \mathbf{x} = \mathbf{y} + (-\mathbf{x})$ . It should be noted that in this example the real numbers may be replaced by the complex numbers; in that case the corresponding vector space  $U^n$  is referred to as the unitary  $n$ -dimensional space.

Actually  $E^n$  (and  $U^n$ ) are not merely vector spaces; we can perform the additional operation of forming the inner product  $(\mathbf{x}, \mathbf{y})$  of two elements  $\mathbf{x}, \mathbf{y}$ , which has a value

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \bar{y}_i,$$

where  $\bar{y}_i$  is the complex conjugate of  $y_i$ . (Two elements  $\mathbf{x}, \mathbf{y}$  are said to be *orthogonal* if  $(\mathbf{x}, \mathbf{y}) = 0$ .) By means of this operation we define the usual distance function

$$d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y})^{\frac{1}{2}} = \left( \sum_{i=1}^n |x_i - y_i|^2 \right)^{\frac{1}{2}},$$

where  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \geq 0$  (with equality holding if and only if  $\mathbf{x} = \mathbf{y}$ ), and the triangle inequality

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}).$$

Conventionally,  $x_i$  is called the *i*th *component* of  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , and  $d(\mathbf{0}, \mathbf{x})$  is called the *length* of  $\mathbf{x}$ . In all that follows we shall be considering  $E^n$  and  $U^n$  as the vector spaces involved, and any statements made concerning  $E^n$  may be applied with equal force to  $U^n$  (but not conversely). To eliminate reference to real and complex numbers we shall refer to the numbers involved as *scalars* ( $\alpha, \beta, \gamma$ , etc.), and to the elements of a vector space as *vectors* ( $\mathbf{x}, \mathbf{y}, \mathbf{z}$ , etc.).

### EXERCISES

1. Let  $\mathcal{P}^n$  be the set of all polynomials  $p(t)$  defined on  $-\infty < a \leq t \leq b < \infty$ , with real (complex) coefficients such that the degree of  $p(t)$  is at most  $n - 1$ . Verify that under the ordinary notions of addition of polynomials and multiplication by a real (complex) scalar,  $\mathcal{P}^n$  is a real (complex) vector space. Show that an inner product for  $p_1$  and  $p_2$  in  $\mathcal{P}^n$  may be defined by

$$(p_1, p_2) = \int_a^b p_1(t) \bar{p}_2(t) dt.$$

2. Let  $C$  be the set of complex numbers, addition and multiplication by *real* scalars being defined as usual. Show that  $C$  is a real vector space, which possesses an inner product for  $x = a + bi$  and  $y = c + di$  given by  $(x, y) = ac + bd$ .

3. In  $E^3$ , for instance, we can associate each vector  $\langle x, y, z \rangle$  with the point in three-dimensional space having Cartesian coordinates  $\{x, y, z\}$ . Thus in referring to sets of vectors we may speak of points, lines, and planes. Interpret this statement.

4. For  $E^n$  let

$$d(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq n} |x_i - y_i|.$$

Show that  $d(\mathbf{x}, \mathbf{y})$  satisfies the usual requirements for a distance function; namely,

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= d(\mathbf{y}, \mathbf{x}) \geq 0 \quad (\text{with equality if and only if } \mathbf{x} = \mathbf{y}), \\ d(\mathbf{x}, \mathbf{y}) &\leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}). \end{aligned}$$

5. Using the usual distance function for  $E^n$  and noticing the relationships between "orthogonal" and the usual definition of "perpendicular," state and prove a general Pythagorean theorem for  $E^n$ .

**A.2 Subspaces, linear independence, basis, direct sums, orthogonal complements.** A subspace  $V_1$  of a vector space  $V$  is a subset which is closed under addition and scalar multiplication: if  $\mathbf{x}$  and  $\mathbf{y}$  are in  $V_1$ , and  $\alpha$  is a scalar, then  $\mathbf{x} + \mathbf{y}$  and  $\alpha\mathbf{x}$  are in  $V_1$ . Under these operations  $V_1$  is a vector space in its own right. For example, the set of all linear combinations

$$\sum_{i=1}^k \alpha_i \mathbf{x}^i = \alpha_1 \mathbf{x}^1 + \cdots + \alpha_k \mathbf{x}^k$$

of a fixed set  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of *vectors* forms a subspace called the span of  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$ . Alternatively, if  $V_1$  is the span of  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$ , then  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$  spans  $V_1$ .

A subset  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$  is *linearly independent* if

$$\sum_{i=1}^k \alpha_i \mathbf{x}^i = \mathbf{0}$$

implies  $\alpha_i = 0$  ( $i = 1, \dots, k$ ). A *linearly independent* subset  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$  which spans  $V$  is a *basis* for  $V$ . In  $E^n$ ,

$$\begin{aligned} \mathbf{e}^1 &= \langle 1, 0, 0, \dots, 0 \rangle, & \mathbf{e}^2 &= \langle 0, 1, 0, \dots, 0 \rangle, & \dots, \\ \mathbf{e}^n &= \langle 0, 0, \dots, 0, 1 \rangle \end{aligned}$$

clearly forms a basis. (Not all vector spaces have finite bases; a trivial example is the case in which  $V$  has exactly one element,  $\mathbf{0}$ , where no subset is independent. In this case we shall not usually regard  $V$  as a vector space.)

If  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$  spans  $V$ , then a subset forms a basis for  $V$ . This is clear if the set is linearly independent; otherwise we have

$$\sum_{i=1}^k \alpha_i \mathbf{x}^i = \mathbf{0},$$

with, say,  $\alpha_1 \neq 0$ . It follows that

$$\mathbf{x}^1 = \sum_{i=2}^k (-\alpha_i \cdot \alpha_1^{-1}) \mathbf{x}^i,$$

and clearly  $\{\mathbf{x}^2, \dots, \mathbf{x}^k\}$  spans  $V$ . If this subset is not independent, we repeat the argument until we arrive at one that is. The advantage of recognizing a basis  $\{\mathbf{y}^1, \dots, \mathbf{y}^l\}$  for  $V$  is that each  $\mathbf{x}$  in  $V$  is *uniquely* expressible in the form

$$\sum_{i=1}^l \alpha_i \mathbf{y}^i,$$

where we refer to the unique scalars  $\alpha_i$  as the coordinates of  $\mathbf{x}$  relative to the basis  $\{\mathbf{y}_1, \dots, \mathbf{y}_l\}$ .

Some important facts about bases are the following, which we state without proof. Each nontrivial subspace  $V_1$  of  $E^n$  has a basis, and each basis for  $V_1$  has the same number of elements, the *dimension* (abbreviated *dim*) of  $V_1$ . If  $V_1 = \{\mathbf{0}\}$ , then  $\dim V_1 = 0$ . If  $V_1 \neq E^n$  (i.e.,  $V_1$  is a *proper* subspace), then each basis  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  of  $V_1$  can be enlarged to a basis  $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k, \mathbf{y}^1, \dots, \mathbf{y}^l\}$  of  $E^n$  (thus  $V_1$  has dimension  $k < n = \dim E^n$ ), and indeed any linearly independent subset can

be so enlarged, being a basis of its span. Consequently, any linearly independent subset of  $E^n$  with  $n$  elements is already a basis for  $E^n$ . It should be noted that  $V_1$  as a vector space is essentially  $E^k$  imbedded in  $E^n$ , since the one-to-one (1-1) mapping

$$\langle \alpha_1, \alpha_2, \dots, \alpha_k \rangle \rightarrow \sum_{i=1}^k \alpha_i \mathbf{x}^i$$

( $\alpha_i$  scalars) of  $E^k$  onto  $V_1$  preserves all sums and scalar multiples, as does its inverse.

In  $E^n$  or  $U^n$  a basis is *orthonormal* if

$$(\mathbf{y}^i, \mathbf{y}^j) = \begin{cases} 0 & i \neq j, \\ 1 & i = j, \end{cases}$$

i.e., if the distinct elements of the basis are orthogonal and each is of length 1. The components of a vector relative to an orthonormal basis are computed from the inner product:

$$\mathbf{x} = \sum_{i=1}^k (\mathbf{x}, \mathbf{y}^i) \mathbf{y}^i$$

for  $\mathbf{x}$  in the span of  $\{\mathbf{y}^1, \dots, \mathbf{y}^k\}$ , since the difference of the two sides is orthogonal to each  $\mathbf{y}^i$ , and hence to itself. *Every subspace of  $E^n$  (or  $U^n$ ) has an orthonormal basis.*

If  $V_1$  and  $V_2$  are subspaces of  $V$  and

$$V = V_1 + V_2 = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in V_1, \mathbf{y} \in V_2\}$$

while  $V_1 \cap V_2 = \{\mathbf{0}\}$ , then  $V$  is the *direct sum* of  $V_1$  and  $V_2$ ; symbolically  $V = V_1 \oplus V_2$ . Each  $\mathbf{z}$  in  $V$  is then uniquely expressible as  $\mathbf{x} + \mathbf{y}$ , where  $\mathbf{x} \in V_1$ ,  $\mathbf{y} \in V_2$ ; for  $\mathbf{x} + \mathbf{y} = \mathbf{x}' + \mathbf{y}'$  implies that  $\mathbf{x} - \mathbf{x}' = \mathbf{y}' - \mathbf{y}$  is in both  $V_1$  and  $V_2$ , and hence equals zero. Consequently, if  $\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$  is a basis for  $V_1$  and  $\{\mathbf{y}^1, \dots, \mathbf{y}^m\}$  is a basis for  $V_2$ , then

$$\{\mathbf{x}^1, \dots, \mathbf{x}^k, \mathbf{y}^1, \dots, \mathbf{y}^m\}$$

is a basis for  $V$ ; for it clearly spans  $V$ , while  $\sum \gamma_i \mathbf{x}^i + \sum \delta_j \mathbf{y}^j = \mathbf{0}$  or  $\sum \gamma_i \mathbf{x}^i = -\sum \delta_j \mathbf{y}^j$  implies that each is zero, and hence that all  $\gamma_i$  and  $\delta_j$  are 0.

For any subspace  $V$  of  $E^n$  (or  $U^n$ ) the set

$$V^\perp = \{\mathbf{x} \mid (\mathbf{x}, \mathbf{y}) = 0 \quad \text{for } \mathbf{y} \text{ in } V\}$$

is called the *orthogonal complement* of  $V$ . It is clearly a subspace, and  $E^n = V \oplus V^\perp$ ; also  $V^{\perp\perp} = (V^\perp)^\perp = V$ . Since clearly  $\dim(V_1 \oplus V_2) = \dim V_1 + \dim V_2$ , it follows that  $\dim V^\perp = n - \dim V$ .

## EXERCISES

1. In  $P^n$  (see Problem 1 of Section A.1) show that the polynomials  $1, t, \dots, t^{n-1}$  comprise a basis, and that  $P^n$  has dimension  $n$ .
2. With  $C$  as specified in Problem 2 of Section A.1, prove that an orthogonal basis is formed from the elements 1 and  $i$ , and thus that  $C$  is two-dimensional.
3. Prove that any nonvoid proper subspace of  $E^3$  corresponds to a line or plane passing through the origin, and that its orthogonal complement corresponds to the perpendicular plane or line also passing through the origin.
4. Let  $\mathbf{a}$  be a nonzero vector in  $E^n$ . Let  $H = \{\mathbf{x} | \mathbf{x} \in E^n, (\mathbf{x}, \mathbf{a}) = 0\}$ . Show that  $H$ , called a hyperplane, is a subspace of  $E^n$  and find its dimension.
5. Find an orthogonal basis for  $P^2$  and express the polynomials 1 and  $t$  in terms of this basis.

**A.3 Linear transformations, matrices, and linear equations.** A linear transformation  $T$  mapping  $E^n$  into  $E^m$  is a function with domain  $E^n$  and values in  $E^m$  which preserves sums and scalar multiples:  $T(\mathbf{x} + \mathbf{y}) = T\mathbf{x} + T\mathbf{y}$ , and  $T(\alpha\mathbf{x}) = \alpha T\mathbf{x}$ . Clearly,  $T\mathbf{0} = \mathbf{0}$ , since  $T\mathbf{0} = T(\mathbf{0} + \mathbf{0}) = T\mathbf{0} + T\mathbf{0}$ ; similarly,  $T(\mathbf{x} - \mathbf{y}) = T\mathbf{x} - T\mathbf{y}$ , since

$$T(\mathbf{x} - \mathbf{y}) + T\mathbf{y} = T(\mathbf{x} - \mathbf{y} + \mathbf{y}) = T\mathbf{x}.$$

The null space of  $T$ ,  $\mathcal{N}(T) = \{\mathbf{x} | T\mathbf{x} = \mathbf{0}\}$ , is a subspace of the domain  $E^n$  (that is,  $\mathbf{x}, \mathbf{y}$  in  $\mathcal{N}(T)$  yields  $\alpha\mathbf{x} + \beta\mathbf{y}$  in  $\mathcal{N}(T)$  for any scalars  $\alpha$  and  $\beta$ ); however,  $\mathcal{N}(T)$  can reduce to the trivial subspace  $\{\mathbf{0}\}$ . This occurs exactly when  $T$  is one-to-one, and  $T$  being one-to-one is equivalent to  $\mathcal{N}(T) = \{\mathbf{0}\}$ , or  $\dim \mathcal{N}(T) = 0$ .

Similarly, the *range* of  $T$ ,  $\mathcal{R}(T) = \{T\mathbf{x} | \mathbf{x} \in E^n\}$ , is a subspace of  $E^m$ , and the following dimensional relationship holds:

$$\dim \mathcal{N}(T) + \dim \mathcal{R}(T) = n. \quad (\text{A.3.1})$$

We demonstrate this relationship as follows. If we select a basis

$$\{\mathbf{x}^1, \dots, \mathbf{x}^k\}$$

of  $\mathcal{N}(T)$  and enlarge it to a basis  $\{\mathbf{x}^1, \dots, \mathbf{x}^k, \mathbf{y}^{k+1}, \dots, \mathbf{y}^n\}$  of  $E^n$ , then, with  $V = \text{span}\{\mathbf{y}^{k+1}, \dots, \mathbf{y}^n\}$ , we have  $E^n = \mathcal{N}(T) \oplus V$ . We have only to check  $\mathcal{N}(T) \cap V = \{\mathbf{0}\}$ . But  $\mathbf{z}$  in  $\mathcal{N}(T) \cap V$  implies

$$\mathbf{z} = \sum_{i=1}^k \alpha_i \mathbf{x}^i = \sum_{j=k+1}^n \beta_j \mathbf{y}^j;$$

hence all  $\alpha_i, \beta_j = 0$ , since  $\{\mathbf{x}^1, \dots, \mathbf{x}^k, \mathbf{y}^{k+1}, \dots, \mathbf{y}^n\}$  is linearly independent, and thus  $\mathbf{z} = \mathbf{0}$ . But the image of the general element  $\sum \alpha_i \mathbf{x}^i + \sum \beta_j \mathbf{y}^j$  of  $E^n$  is  $T(\sum \beta_j \mathbf{y}^j) = \sum \beta_j T\mathbf{y}^j$ , and hence  $\{T\mathbf{y}^{k+1}, \dots, T\mathbf{y}^n\}$  spans

$\mathcal{R}(\mathbf{T})$ . On the other hand, the set is linearly independent, since  $\sum \beta_j \mathbf{T} \mathbf{y}^j = \mathbf{0}$  implies that  $\mathbf{T}(\sum \beta_j \mathbf{y}^j) = \mathbf{0}$ , which means that  $\sum \beta_j \mathbf{y}^j \in \mathcal{N}(\mathbf{T}) \cap V = \{\mathbf{0}\}$ , and hence all  $\beta_j = 0$ . Thus  $\dim \mathcal{R}(\mathbf{T}) = n - k$ , and our relation holds. It should be noted that  $V$  is in no way unique.

As consequences, note that  $\dim \mathcal{R}(\mathbf{T}) \leq n$ , with equality holding if and only if  $\mathbf{T}$  is 1-1. Thus when  $m = n$ , the transformation  $\mathbf{T}$  maps  $E^n$  onto itself if and only if it is 1-1.

*Matrices.* A rectangular array of numbers

$$\mathbf{A} = \|a_{ij}\| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}$$

is called a *matrix*—in this case an  $m \times n$  ( $m$ -row by  $n$ -column) matrix—with *entries*  $a_{ij}$ . If  $m = n$ , we call the matrix *square*.

$\mathbf{A}$  determines a linear transformation  $\mathbf{T}$  of  $E^n$  into  $E^m$  defined by setting

$$\mathbf{T}(x_1, x_2, \dots, x_n) = \left\langle \sum_{j=1}^n a_{1j} x_j, \quad \sum_{j=1}^n a_{2j} x_j, \dots, \quad \sum_{j=1}^n a_{mj} x_j \right\rangle.$$

Conversely, if we denote the usual orthonormal basis of  $E^n$  by  $\{\mathbf{e}^1, \dots, \mathbf{e}^n\}$  and that of  $E^m$  by  $\{\mathbf{f}^1, \dots, \mathbf{f}^m\}$ , each linear transformation  $\mathbf{T}$  is so determined. For  $\mathbf{T}$  is determined by its values at the  $\mathbf{e}^j$ , since

$$\mathbf{T} \left( \sum_{j=1}^n x_j \mathbf{e}^j \right) = \sum_{j=1}^n x_j \mathbf{T} \mathbf{e}^j;$$

and since each  $\mathbf{T} \mathbf{e}^j$  is expressible as

$$\sum_{i=1}^m a_{ij} \mathbf{f}^i$$

for some unique  $a_{1j}, a_{2j}, \dots, a_{mj}$ , we obtain

$$\mathbf{T} \mathbf{x} = \sum_{j=1}^n x_j \left( \sum_{i=1}^m a_{ij} \mathbf{f}^i \right) = \sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} \mathbf{f}^i = \sum_{i=1}^m \left( \sum_{j=1}^n x_j a_{ij} \right) \mathbf{f}^i,$$

and the components of  $\mathbf{T} \mathbf{x}$  are obtained as before from  $\mathbf{A}$ . Moreover, it is clear that we may replace  $\{\mathbf{e}^1, \dots, \mathbf{e}^n\}$  and  $\{\mathbf{f}^1, \dots, \mathbf{f}^m\}$  by any other fixed ordered bases in  $E^n$  and  $E^m$  and set up a similar one-to-one correspondence between linear transformations mapping  $E^n$  into  $E^m$  and  $m \times n$  matrices. The exact correspondence will of course depend greatly on the bases chosen (and on their indexing). However, when we consider a matrix as a transformation we shall have in mind our original correspondence.

Corresponding to the linear transformations  $\mathbf{T}$  and  $\mathbf{S}$  mapping  $E^n$  into  $E^m$  there is a natural sum  $\mathbf{T} + \mathbf{S}$ , again a linear transformation, such that  $(\mathbf{T} + \mathbf{S})(\mathbf{x}) = \mathbf{Tx} + \mathbf{Sx}$ . Similarly,  $c\mathbf{T}$  at  $\mathbf{x}$  is defined to be  $c \cdot \mathbf{Tx}$  ( $c$  a scalar). The corresponding operations on matrices are clearly the addition of corresponding entries:

$$\mathbf{A} + \mathbf{B} = \begin{vmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots \\ a_{21} + b_{21} & \dots & \dots \\ \vdots & & \\ a_{m1} + b_{m1} & \dots & \dots \end{vmatrix}$$

and

$$c\mathbf{A} = \begin{vmatrix} ca_{11} & ca_{12} & \dots \\ \vdots & \vdots & \\ ca_{m1} & ca_{m2} & \dots \end{vmatrix}$$

(each entry multiplied by  $c$ ). Clearly the  $m \times n$  matrices thus form a vector space with an obvious basis  $\mathbf{e}_{ij}$  ( $i = 1, \dots, m$ ;  $j = 1, \dots, n$ ), where  $\mathbf{e}_{ij}$  has all entries 0 except for a single 1 in the  $ij$ th place. Thus the space is  $mn$ -dimensional.

If  $\mathbf{T}$  maps  $E^n$  into  $E^m$  and  $\mathbf{S}$  maps  $E^m$  into  $E^l$ , we can compose the mappings in such a way that  $\mathbf{ST}$  maps  $E^n$  into  $E^l$ . Explicitly, we set  $\mathbf{ST}(\mathbf{x}) = \mathbf{S}(\mathbf{Tx})$ . The linear transformation  $\mathbf{ST}$  is called the product of  $\mathbf{S}$  and  $\mathbf{T}$ ; the corresponding matrix product is easily computed: with  $\mathbf{S}$  corresponding to  $\mathbf{A}(l \times m)$  and  $\mathbf{T}$  corresponding to  $\mathbf{B}(m \times n)$ ,  $\mathbf{AB}$  is an  $l \times n$  matrix with  $ij$ th entry

$$\sum_{k=1}^m a_{ik}b_{kj} \quad (i = 1, \dots, l; \quad j = 1, \dots, n).$$

If we view the elements of  $E^n$  as  $n \times 1$  matrices

$$\begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{vmatrix}$$

(a *column* vector), our computation of the components of  $\mathbf{Tx}$  (where  $\mathbf{T}$  corresponds to  $\|b_{ij}\|$ ) is simply the formation of the matrix product

$$\begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{vmatrix} = \begin{vmatrix} \sum_{j=1}^k b_{1j}x_j \\ \vdots \\ \sum_{j=1}^n b_{mj}x_j \end{vmatrix}$$

giving the image elements of  $E^m$  as column vectors ( $m \times 1$ ). The resulting column vector can also be expressed as

$$x_1 \begin{vmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{m1} \end{vmatrix} + x_2 \begin{vmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{m2} \end{vmatrix} + \cdots + x_n \begin{vmatrix} b_{1n} \\ b_{2n} \\ \vdots \\ b_{mn} \end{vmatrix}$$

so that the *dimension of  $\mathcal{R}(\mathbf{T})$ , called the rank of  $\mathbf{T}$  or  $\mathbf{B}$ , is the number of linearly independent columns of  $\mathbf{B}$*  (viewed as vector elements of  $E^m$ ). One can obtain the rank by eliminating dependent columns, or by determinants; the rank is  $k$  if after deleting rows and columns of  $\mathbf{B}$ , the largest square submatrices with nonzero determinant are  $k \times k$ .

Finally, a linear transformation from  $E^n$  into  $E^1$ , which we may regard as the real line, is called a *linear functional*. As we have seen, it is given (relative to the bases  $\{\mathbf{e}^1, \dots, \mathbf{e}^n\}$  and  $\{\mathbf{f}^1\}$ , where  $\mathbf{f}^1 = \langle 1 \rangle$ ) by a  $1 \times n$  matrix, and its value at  $\langle x_1, x_2, \dots, x_n \rangle$  is of the form  $c_1x_1 + c_2x_2 + \cdots + c_nx_n$ , with  $c_i$  fixed. Thus each linear functional is obtained from the inner product and maps  $\mathbf{x}$  into  $\langle \mathbf{x}, \mathbf{y} \rangle$  for some fixed  $\mathbf{y}$  in  $E^n$ .

*Linear equations.* The correspondence between matrices and linear transformations can be applied to systems of linear equations. Consider the following system of  $m$  equations in  $n$  unknown (real or complex) numbers  $x_1, x_2, \dots, x_n$ :

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m. \end{aligned} \tag{A.3.2}$$

Solving the system is equivalent to finding an  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  in  $E^n$  mapping onto  $\mathbf{c} = \langle c_1, c_2, \dots, c_m \rangle$ . A solution can be obtained for every  $\mathbf{c}$  if and only if the rank of  $\mathbf{A} = \|a_{ij}\|$  is  $m$ .

If each  $c_i = 0$ , the system is called *homogeneous* (otherwise inhomogeneous) and has the trivial solution  $\langle 0, 0, \dots, 0 \rangle$ ; nontrivial solutions exist whenever  $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$ . If  $n > m$  (more unknowns than equations), this must be the case, for otherwise  $\dim \mathcal{R}(\mathbf{A}) = n$ , which is greater than the dimension of  $E^m$ , while  $\mathcal{R}(\mathbf{A}) \subset E^m$ . If  $n = m$ , (A.3.2) is solvable for each  $\mathbf{c}$  if and only if each solution is unique, or indeed if the solution for  $\mathbf{c} = \mathbf{0}$  is unique; this amounts to saying that  $\mathbf{A}$  maps  $E^n$  onto itself if and only if  $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$ .

*Transformations from  $E^n$  into  $E^n$ .* Let  $L(E^n)$  be the set of all linear transformations  $\mathbf{T}$  mapping  $E^n$  into  $E^n$ . If  $\mathbf{S}$  and  $\mathbf{T}$  are in  $L(E^n)$ , then  $\mathbf{S} + \mathbf{T}$ ,  $\mathbf{ST}$ , and  $c\mathbf{T}$  are all elements of  $L(E^n)$ , and  $L(E^n)$  forms a vector

space with a (noncommutative) associative and distributive multiplication. The zero element maps each  $\mathbf{x}$  into  $\mathbf{0}$ , while the identity transformation  $\mathbf{I}$  mapping  $\mathbf{x}$  into  $\mathbf{x}$  yields  $\mathbf{T}\mathbf{I} = \mathbf{I}\mathbf{T} = \mathbf{T}$ . In terms of matrices, the zero transformation corresponds to the matrix with all entries 0, and

$$\mathbf{I} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}$$

(with 1's only on the main diagonal). We know that  $\mathbf{T}$  is 1-1 if and only if  $\mathbf{T}$  is onto; in this case the inverse transformation  $\mathbf{T}^{-1}$  (which has as its domain all of  $E^n$ ) is also linear and thus in  $L(E^n)$ ;  $\mathbf{T}^{-1}\mathbf{y} = \mathbf{x}$  is equivalent by definition to  $\mathbf{y} = \mathbf{T}\mathbf{x}$ . The 1-1 transformations in  $L(E^n)$  are just those with a multiplicative inverse  $\mathbf{T}^{-1}$ :  $\mathbf{T}\mathbf{T}^{-1} = \mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$ . They are called *nonsingular* or invertible; all others are *singular*. The same terminology is applied to matrices. In particular,  $\mathbf{A}$  is nonsingular if  $\mathbf{AB} = \mathbf{I}$  (or  $\mathbf{BA} = \mathbf{I}$ ) for some matrix  $\mathbf{B}$  (for  $\mathbf{A}$ , considered as a linear transformation, cannot lower dimension, and therefore is onto and 1-1).  $\mathbf{B}$  of course must be the (unique) inverse matrix, denoted by  $\mathbf{A}^{-1}$ , and

$$\sum_{j=1}^n a_{ij}b_{jk} = 0$$

if  $i \neq k$ . If  $\mathbf{A}$  is singular, we have numbers  $x_1, x_2, \dots, x_n$ , not all zero, with  $\mathbf{A}\langle x_1, \dots, x_n \rangle = \langle 0, \dots, 0 \rangle$ .

Finally note that a product of nonsingular matrices or transformation is nonsingular:  $(\mathbf{ST}) \cdot (\mathbf{T}^{-1}\mathbf{S}^{-1}) = \mathbf{T}^{-1}\mathbf{S}^{-1} \cdot \mathbf{ST} = \mathbf{I}$ .

### EXERCISES

1. In  $E^2$  with basis  $\{\mathbf{e}^1, \mathbf{e}^2\}$  let the transformation  $\mathbf{A}_\theta$  describe the rotation of all points  $\theta$  degrees counterclockwise about the origin. Show that the matrix of  $\mathbf{A}_\theta$  is given by

$$\mathbf{A}_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

Show further that  $\mathbf{A}_{\theta_1 + \theta_2} = \mathbf{A}_{\theta_1}\mathbf{A}_{\theta_2}$  and  $\mathbf{A}_\theta^{-1} = \mathbf{A}_{-\theta}$ .

2. In the space  $\mathcal{P}^n$  show that

$$\int_a^b p(t) dt$$

is a linear functional and express it in the form  $(\mathbf{q}, \mathbf{p})$ , where  $\mathbf{q}$  is some element in  $\mathcal{P}^n$  and  $(,)$  is the inner product specified in Problem 1 of Section A.1.

3. On the space  $P^n$  with the basis  $\{1, t, t^2, \dots, t^{n-1}\}$ , define the linear transformation  $\mathbf{A}_k$  ( $k \geq 1$ ) as follows:

$$\mathbf{A}_k \mathbf{p} = \frac{d^k}{dt^k} p(t).$$

Verify that  $\mathbf{A}_k$  is a singular transformation whose matrix representation in terms of this basis is

$$\left( \begin{array}{cccccc|c} 0 & \dots & 0 & \frac{k!}{0!} & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \frac{(k+1)!}{1!} & 0 & \dots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \frac{(n-1)!}{(n-k-1)!} \\ \vdots & & & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \end{array} \right)$$

Show also that  $\mathcal{R}(\mathbf{A}_k)$  is the span of  $\{t^{n-k}, t^{n-k+1}, \dots, t^{n-1}\}$ , and that  $\mathcal{N}(\mathbf{A}_k)$  is the span of  $\{1, t, \dots, t^{n-k-1}\}$ .

**A.4 Eigenvalues, eigenvectors, and the Jordan canonical form.** If  $\mathbf{T} - \lambda \mathbf{I}$  is singular for some scalar  $\lambda$ , then  $\lambda$  is called an *eigenvalue* of  $\mathbf{T}$ . Associated with  $\lambda$  we have some  $\mathbf{x} \neq \mathbf{0}$  such that  $(\mathbf{T} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$  or  $\mathbf{T}\mathbf{x} = \lambda\mathbf{x}$ ; this vector  $\mathbf{x}$  is called an *eigenvector*, corresponding to the eigenvalue  $\lambda$ .

Each  $\mathbf{T}$  in  $L(U^n)$  has eigenvalues, and hence eigenvectors. The corresponding statement for  $L(E^n)$ , where  $\lambda$  is taken to be real, is false. However, any real matrix corresponds to an element of  $L(U^n)$  and thus has complex eigenvalues and eigenvectors.

The existence of eigenvalues is established as follows. Note that since  $L(U^n)$  is  $n^2$ -dimensional, the  $n^2 + 1$  elements  $\mathbf{I}, \mathbf{T}, \mathbf{T}^2, \dots, \mathbf{T}^{n^2}$  are linearly dependent; we have  $a_0, a_1, \dots, a_{n^2}$  not all zero, with  $a_0\mathbf{I} + a_1\mathbf{T} + \dots + a_{n^2}\mathbf{T}^{n^2} = \mathbf{0}$ . The corresponding polynomial  $p(z) = a_0 + a_1z + \dots + a_{n^2}z^{n^2}$  can of course be factored by the fundamental theorem of algebra:

$$p(z) = \prod_{i=1}^k (z - \lambda_i).$$

Since the identity of the polynomials says exactly that we obtain the coefficients of  $p$  by multiplying out the right side, and the same coefficients are obtained in multiplying out

$$\prod_{i=1}^k (\mathbf{T} - \lambda_i \mathbf{I}),$$

we have

$$\prod_{i=1}^k (\mathbf{T} - \lambda_i \mathbf{I}) = a_0 \mathbf{I} + a_1 \mathbf{T} + \cdots + a_{n^2} \mathbf{T}^{n^2} = p(\mathbf{T}) = 0;$$

consequently one of the  $\mathbf{T} - \lambda_i \mathbf{I}$  is singular.

For any basis  $\{\mathbf{y}^1, \dots, \mathbf{y}^n\}$  of  $U^n$ , by writing

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{y}^i$$

and similarly expressing  $\mathbf{T}\mathbf{x}$  in terms of its components relative to  $\{\mathbf{y}^1, \dots, \mathbf{y}^n\}$ , we can represent  $\mathbf{T}$  by a matrix  $\mathbf{A}$  mapping  $\langle x_1, \dots, x_n \rangle$  into the vector of components of  $\mathbf{T}\mathbf{x}$  relative to the basis. For an appropriate choice of basis,  $\mathbf{A}$  has the Jordan *canonical form*

$$\mathbf{A} = \begin{vmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k \end{vmatrix}$$

where  $\mathbf{0}$  represents appropriate rectangular matrices with all entries zero, and each  $\mathbf{A}_i$  is a square matrix which can similarly be written (in so-called block-diagonal form) with diagonal blocks of the form

$$\mathbf{B} = \begin{vmatrix} \lambda_i & 0 & 0 & \cdots & 0 \\ 1 & \lambda_i & 0 & \cdots & 0 \\ 0 & 1 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{vmatrix}$$

( $\lambda_i$  on the main diagonal, 1's below, 0's elsewhere). The same  $\lambda_i$  appears in all sub-blocks of  $\mathbf{A}_i$ , but  $\lambda_i \neq \lambda_j$ ,  $i \neq j$ , and  $\lambda_1, \dots, \lambda_k$  comprise the set of all eigenvalues of  $\mathbf{T}$ . The effect of this decomposition of  $\mathbf{A}$  into blocks is the following: Any vector in the span of the basis elements

$$\{\mathbf{y}^j, \mathbf{y}^{j+1}, \dots, \mathbf{y}^k\}$$

corresponding to the block  $\mathbf{B}$  maps into a vector in this same subspace; moreover, on this subspace  $\mathbf{T}$  acts as the sum of two simple transformations: the transformation that maps  $\mathbf{x}$  into  $\lambda_i \mathbf{x}$  and the transformation that maps  $\mathbf{y}^j$  into  $\mathbf{y}^{j+1}$ ,  $\mathbf{y}^{j+1}$  into  $\mathbf{y}^{j+2}, \dots, \mathbf{y}^{k-1}$  into  $\mathbf{y}^k$ ,  $\mathbf{y}^k$  into  $\mathbf{0}$  (thus  $\mathbf{T} - \lambda_i \mathbf{I}$  reduces to the latter transformation on the subspace). In effect, we have a decomposition of  $E^n$  into (a direct sum of) pieces, and  $\mathbf{T}$  has a simple form on each piece. In many important cases each block  $\mathbf{B}$  is  $1 \times 1$ ; the matrix  $\mathbf{A}$  assumes the pure diagonal form

$$\begin{vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{vmatrix},$$

where now the  $\lambda_i$  need not all be distinct; and the eigenvectors of  $\mathbf{T}$  span  $U^n$ . Moreover, if  $\mathbf{T}$  has  $n$  distinct eigenvalues, it is clear that each block  $\mathbf{B}$  is  $1 \times 1$  and  $\mathbf{A}$  has the pure diagonal form.

We now indicate how the basis  $\{\mathbf{y}^1, \dots, \mathbf{y}^n\}$  is constructed. For a singular  $\mathbf{S}$  in  $L(U^n)$ , or indeed  $L(E^n)$ , consider  $\mathfrak{N}(\mathbf{S})$ ,  $\mathfrak{N}(\mathbf{S}^2)$ ,  $\dots$ . Clearly,  $\mathfrak{N}(\mathbf{S}) \subset \mathfrak{N}(\mathbf{S}^2) \subset \dots$ , and since proper subspaces have smaller dimensions, there is some integer  $k$  ( $\leq n$ ) for which

$$\mathfrak{N}(\mathbf{S}) \subsetneq \mathfrak{N}(\mathbf{S}^2) \subsetneq \dots \subsetneq \mathfrak{N}(\mathbf{S}^k) = \mathfrak{N}(\mathbf{S}^{k+1}).$$

Then  $\mathfrak{N}(\mathbf{S}^k) = \mathfrak{N}(\mathbf{S}^{k+j})$  for all  $j \geq 0$ . For since  $\mathfrak{N}(\mathbf{S}^k) = \mathfrak{N}(\mathbf{S}^{k+1})$ , if  $\mathbf{S}^{k+1}\mathbf{x} = \mathbf{0}$  it follows that  $\mathbf{S}^k\mathbf{x} = \mathbf{0}$ ; and thus  $\mathbf{S}^{k+j}\mathbf{x} = \mathbf{0} = \mathbf{S}^{k+1}(\mathbf{S}^{j-1}\mathbf{x})$  implies  $\mathbf{S}^k(\mathbf{S}^{j-1}\mathbf{x}) = \mathbf{S}^{k+j-1}\mathbf{x} = \mathbf{0}$ , from which it follows that

$$\mathbf{S}^{k+j-1}\mathbf{x} = \mathbf{S}^{k+j-2}\mathbf{x} = \dots = \mathbf{S}^k\mathbf{x} = \mathbf{0}.$$

By the dimension relation (A.3.1),

$$\mathfrak{R}(\mathbf{S}) \supsetneq \mathfrak{R}(\mathbf{S}^2) \supsetneq \dots \supsetneq \mathfrak{R}(\mathbf{S}^k) = \mathfrak{R}(\mathbf{S}^{k+1}) = \dots$$

Thus  $\mathbf{S}$  restricted to  $\mathfrak{R}(\mathbf{S}^k)$  is 1-1 because it is onto. But, since  $\mathbf{S}^k$  is 1-1, we also have  $\mathfrak{N}(\mathbf{S}^k) \cap \mathfrak{R}(\mathbf{S}^k) = \{\mathbf{0}\}$ . If  $\{\mathbf{x}^1, \dots, \mathbf{x}^l\}$  is a basis for  $\mathfrak{N}(\mathbf{S}^k)$  and  $\{\mathbf{y}^1, \dots, \mathbf{y}^m\}$  is a basis for  $\mathfrak{R}(\mathbf{S}^k)$ , it follows by (A.3.1) that  $l + m = n$ . The conjunction  $\{\mathbf{x}^1, \dots, \mathbf{x}^l; \mathbf{y}^1, \dots, \mathbf{y}^m\}$  forms a basis for  $U^n$ , and  $U^n = \mathfrak{N}(\mathbf{S}^k) \oplus \mathfrak{R}(\mathbf{S}^k)$ . Corresponding to our basis,  $\mathbf{S}$  would have the matricial structure

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix},$$

where  $\mathbf{A}$  is  $l \times l$  and  $\mathbf{B}$  is  $(n - l) \times (n - l)$ .

Now suppose that  $\mathbf{T}$  is any element of  $L(U^n)$ . We know that there is an eigenvalue  $\lambda_1$ ; hence we may let  $\mathbf{S} = \mathbf{T} - \lambda_1\mathbf{I}$  and let

$$\mathfrak{N}_1 = \mathfrak{N}[(\mathbf{T} - \lambda_1\mathbf{I})^{k_1}],$$

with  $k_1$  the  $k$  obtained for  $\mathbf{S}$  above. Accordingly,  $\mathfrak{R}_1 = \mathfrak{R}[(\mathbf{T} - \lambda_1\mathbf{I})^{k_1}]$ . We have  $U^n = \mathfrak{N}_1 \oplus \mathfrak{R}_1$ , with  $(\mathbf{T} - \lambda_1\mathbf{I})\mathbf{x}$  in  $\mathfrak{N}_1$  or  $\mathfrak{R}_1$  according as  $\mathbf{x}$  is in  $\mathfrak{N}_1$  or  $\mathfrak{R}_1$ . Since  $\lambda_1\mathbf{x}$  is in a given subspace whenever  $\mathbf{x}$  is,  $\mathbf{T}\mathbf{x}$  is in  $\mathfrak{N}_1$  if  $\mathbf{x}$  is in  $\mathfrak{N}_1$ , and  $\mathbf{T}\mathbf{x}$  is in  $\mathfrak{R}_1$  if  $\mathbf{x}$  is in  $\mathfrak{R}_1$ . Let  $\mathbf{T}_1$  be the restriction of  $\mathbf{T}$  to

$\mathfrak{R}_1$ , mapping  $\mathfrak{R}_1$  into itself. Since we can identify  $\mathfrak{R}_1$  with some  $U^m$ , it follows that  $\mathbf{T}_1$  has an eigenvalue  $\lambda_2$  such that  $\mathbf{T}\mathbf{x} = \mathbf{T}_1\mathbf{x} = \lambda_2\mathbf{x}$  for  $\mathbf{x} \neq \mathbf{0}$  in  $\mathfrak{R}_1$ . Since  $(\mathbf{T} - \lambda_1\mathbf{I})\mathbf{x} \neq \mathbf{0}$  ( $\mathbf{T} - \lambda_1\mathbf{I}$  being 1-1 on  $\mathfrak{R}_1$ ), we deduce that  $\lambda_2 \neq \lambda_1$ . Defining  $\mathfrak{N}_2$  and  $\mathfrak{R}_2$  in the obvious fashion, we obtain  $\mathfrak{R}_1 = \mathfrak{N}_2 \oplus \mathfrak{R}_2$  (hence  $U^n = \mathfrak{N}_1 \oplus \mathfrak{N}_2 \oplus \mathfrak{R}_2$ ), with

$$\mathbf{T}\mathfrak{N}_2 = \mathbf{T}_1\mathfrak{N}_2 = \{\mathbf{T}_1\mathbf{x} \mid \mathbf{x} \in \mathfrak{N}_2\} \subset \mathfrak{N}_2, \quad \mathbf{T}\mathfrak{R}_2 \subset \mathfrak{R}_2.$$

We continue and obtain  $U^n = \mathfrak{N}_1 \oplus \cdots \oplus \mathfrak{N}_r$ , with associated distinct eigenvalues  $\lambda_1, \dots, \lambda_r$ , such that  $(\mathbf{T} - \lambda_i\mathbf{I})^{k_i}\mathbf{x} = \mathbf{0}$  for all  $\mathbf{x}$  in  $\mathfrak{N}_i$  and some appropriate  $k_i \geq 1$ . Choosing any basis for  $U^n$  which is formed by combining bases of the  $\mathfrak{N}_i$  now yields a matrix  $\mathbf{A}$  in block-diagonal form, since this amounts to saying  $\mathbf{T}\mathfrak{N}_i \subset \mathfrak{N}_i$ .

The further decomposition into sub-blocks is delicate: we shall only note that if we choose an  $\mathbf{x}_1$  in  $\mathfrak{N}_1$  for which  $(\mathbf{T} - \lambda_1\mathbf{I})^{k_1-1}\mathbf{x}_1 \neq \mathbf{0}$  (i.e., an element of  $\mathfrak{N}_1(\mathbf{T} - \lambda_1\mathbf{I})^{k_1}$  not in  $\mathfrak{N}_1(\mathbf{T} - \lambda_1\mathbf{I})^{k_1-1}$ ), a choice that is made possible by the meaning of  $k_1$ , then for  $\mathbf{S} = \mathbf{T} - \lambda_1\mathbf{I}$  the elements  $\mathbf{x}_1, \mathbf{S}\mathbf{x}_1, \dots, \mathbf{S}^{k_1-1}\mathbf{x}_1$  are independent: for if

$$\gamma_1\mathbf{x}_1 + \gamma_2\mathbf{S}\mathbf{x}_1 + \cdots + \gamma_{k_1-1}\mathbf{S}^{k_1-1}\mathbf{x}_1 = \mathbf{0},$$

then by applying  $\mathbf{S}^{k_1-1}$  we obtain  $\gamma_1\mathbf{S}^{k_1-1}\mathbf{x}_1 = \mathbf{0}$  and hence  $\gamma_1 = 0$ ; we then apply  $\mathbf{S}^{k_1-2}, \mathbf{S}^{k_1-3}, \dots$  successively to deduce  $\gamma_2 = \gamma_3 = \cdots = 0$ . These vectors need not constitute a basis for  $\mathfrak{N}_1$ , but it can be shown that there exists a finite set  $\mathbf{x}_1, \mathbf{x}_2, \dots$  such that  $\mathbf{x}_1, \mathbf{S}\mathbf{x}_1, \dots, \mathbf{S}^{k_1-1}\mathbf{x}_1; \mathbf{x}_2, \mathbf{S}\mathbf{x}_2, \dots, \mathbf{S}^{j_2-1}\mathbf{x}_2; \mathbf{x}_3, \mathbf{S}\mathbf{x}_3, \dots, \mathbf{S}^{j_3-1}\mathbf{x}_3; \dots$  form a basis for  $\mathfrak{N}_1$  where  $\mathbf{S}^{j_2}\mathbf{x}_2 = \mathbf{0}, \mathbf{S}^{j_3}\mathbf{x}_3 = \mathbf{0}, \dots, j_2 \leq k_1, j_3 \leq j_2, \dots$  etc.\* It is easily seen that on the span  $V_0$  of  $\mathbf{x}_1, \mathbf{S}\mathbf{x}_1, \dots, \mathbf{S}^{k_1-1}\mathbf{x}_1$  we can write  $\mathbf{S}$  as a matrix of the form

$$\begin{vmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{vmatrix};$$

hence on  $V_0$

$$\mathbf{T} = \lambda_1\mathbf{I} + \mathbf{S} = \begin{vmatrix} \lambda_1 & 0 & \dots & 0 \\ 1 & \lambda_1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_1 \end{vmatrix}.$$

\* For the sake of preserving the continuity here, the proof of this assertion has been placed at the end of the argument (see p. 376).

Clearly, each such sub-block of our final matrix  $\mathbf{A}$  yields exactly one eigenvector (here  $\mathbf{S}^{k_1-1}\mathbf{x}_1$ ). Similar statements apply to the span of

$$\mathbf{x}_2, \mathbf{S}\mathbf{x}_2, \dots, \mathbf{S}^{j_2-1}\mathbf{x}_2, \text{ etc.}$$

The number of eigenvectors we thus obtain corresponding to  $\lambda_1$  is called the *geometric multiplicity* of  $\lambda_1$ ; it is precisely the dimension of the subspace spanned by the eigenvectors corresponding to  $\lambda_1$ . Since  $\mathbf{T}$  maps

$$\begin{aligned}\mathbf{x}_1 &\rightarrow \lambda_1\mathbf{x}_1 + \mathbf{S}\mathbf{x}_1, \\ \mathbf{S}\mathbf{x}_1 &\rightarrow \lambda_1\mathbf{S}\mathbf{x}_1 + \mathbf{S}^2\mathbf{x}_1, \\ &\vdots \\ \mathbf{S}^{k_1-2}\mathbf{x}_1 &\rightarrow \lambda_1\mathbf{S}^{k_1-2}\mathbf{x}_1 + \mathbf{S}^{k_1-1}\mathbf{x}_1, \\ \mathbf{S}^{k_1-1}\mathbf{x}_1 &\rightarrow \lambda_1\mathbf{S}^{k_1-1}\mathbf{x}_1,\end{aligned}$$

$\mathbf{T} - \lambda\mathbf{I}$  maps

$$\begin{aligned}\mathbf{x}_1 &\rightarrow (\lambda_1 - \lambda)\mathbf{x}_1 + \mathbf{S}\mathbf{x}_1, \\ \mathbf{S}\mathbf{x}_1 &\rightarrow (\lambda_1 - \lambda)\mathbf{S}\mathbf{x}_1 + \mathbf{S}^2\mathbf{x}_1, \\ &\vdots \\ \mathbf{S}^{k_1-2}\mathbf{x}_1 &\rightarrow (\lambda_1 - \lambda)\mathbf{S}^{k_1-2}\mathbf{x}_1 + \mathbf{S}^{k_1-1}\mathbf{x}_1, \\ \mathbf{S}^{k_1-1}\mathbf{x}_1 &\rightarrow (\lambda_1 - \lambda)\mathbf{S}^{k_1-1}\mathbf{x}_1.\end{aligned}$$

Consequently, if  $\lambda \neq \lambda_1$ , it follows that  $\mathcal{R}(\mathbf{T} - \lambda\mathbf{I})$  contains  $\mathbf{S}^{k_1-1}\mathbf{x}_1$ , hence  $\mathbf{S}^{k_1-2}\mathbf{x}_1$ , and thus all the elements spanning  $V_0$ . Therefore  $\mathbf{T} - \lambda\mathbf{I}$  maps  $V_0$  onto itself for any  $\lambda \neq \lambda_i$  and all  $i$ , yielding  $\mathcal{N}_i \subset \mathcal{R}(\mathbf{T} - \lambda\mathbf{I})$ . Hence  $\mathbf{T} - \lambda\mathbf{I}$  is onto and  $(\mathbf{T} - \lambda\mathbf{I})^{-1}$  exists, and thus  $\lambda_1, \dots, \lambda_k$  is the full set of eigenvalues.

Finally, note that if  $\mathbf{S}_0$  is the unique linear transformation mapping  $\mathbf{e}^i$  onto  $\mathbf{y}^i$ —that is, if

$$\mathbf{S}_0 \sum_{i=1}^n \gamma_i \mathbf{e}^i = \sum_{i=1}^n \gamma_i \mathbf{y}^i$$

—then for  $\mathbf{z} = \langle \gamma_1, \gamma_2, \dots, \gamma_n \rangle$  we have

$$\mathbf{T}\mathbf{S}_0\mathbf{z} = \sum_{i=1}^n \left( \sum_{j=1}^n a_{ij} \gamma_j \right) \mathbf{y}^i;$$

and therefore (since  $\mathbf{S}_0$ , being onto, is nonsingular)

$$\mathbf{S}_0^{-1}\mathbf{T}\mathbf{S}_0\mathbf{z} = \sum_{i=1}^n \left( \sum_{j=1}^n a_{ij} \gamma_j \right) \mathbf{e}^i \quad (\mathbf{S}_0^{-1}\mathbf{y}^i = \mathbf{e}^i).$$

Thus if  $\mathbf{T}$  corresponds to the matrix  $\mathbf{B}$  relative to the basis  $\{\mathbf{e}^1, \dots, \mathbf{e}^n\}$ , and  $\mathbf{S}_0, \mathbf{S}_0^{-1}$  to  $\mathbf{C}, \mathbf{C}^{-1}$ , respectively, we have  $\mathbf{C}^{-1}\mathbf{B}\mathbf{C} = \mathbf{A}$ , or equivalently  $\mathbf{B} = \mathbf{C}(\mathbf{C}^{-1}\mathbf{B}\mathbf{C})\mathbf{C}^{-1} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ . Two matrices  $\mathbf{A}$  and  $\mathbf{B}$  for which  $\mathbf{B} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$  for some  $\mathbf{C}$  are called *similar*, and any matrix is therefore similar to one of the (block-diagonal) Jordan canonical form. Any two matrices  $\mathbf{A}$  and  $\mathbf{B}$  are similar if and only if their Jordan canonical forms differ only in the order in which the blocks and sub-blocks appear.

We now present the formal proof referred to in the footnote on p. 374.

(i) Suppose that we have a linearly independent set of vectors of  $\mathfrak{N}_1$  of the form

$$\begin{aligned} \mathbf{x}_1, \quad \mathbf{S}\mathbf{x}_1, \quad \dots, \quad \mathbf{S}^{j_1-1}\mathbf{x}_1; \\ \mathbf{x}_2, \quad \mathbf{S}\mathbf{x}_2, \quad \dots, \quad \mathbf{S}^{j_2-1}\mathbf{x}_2; \\ \vdots \\ \mathbf{x}_r, \quad \mathbf{S}\mathbf{x}_r, \quad \dots, \quad \mathbf{S}^{j_r-1}\mathbf{x}_r, \end{aligned}$$

where  $k_1 = j_1 \geq j_2 \geq \dots \geq j_r$  and  $\mathbf{S}^{j_i}\mathbf{x}_i = \mathbf{0}$  ( $i = 1, 2, \dots, r$ ).

(ii) Suppose also that for each  $l = 1, 2, \dots, r$  there exists no  $\mathbf{x}$  satisfying  $\mathbf{S}^{k_1}\mathbf{x} = \mathbf{0}$  and such that  $\mathbf{x}, \mathbf{S}\mathbf{x}, \dots, \mathbf{S}^{j_l}\mathbf{x}$  together with the first  $l - 1$  rows of vectors in (i) are linearly independent.

Then either the vectors in (i) form a basis for  $\mathfrak{N}_1$  or they can be extended so as to satisfy both (i) and (ii).

*Proof.* Let  $j_{r+1}$  be the largest integer satisfying the properties that for some  $\mathbf{x}_{r+1}$ ,  $\mathbf{S}^{k_1}\mathbf{x}_{r+1} = \mathbf{0}$  and  $\mathbf{x}_{r+1}, \mathbf{S}\mathbf{x}_{r+1}, \dots, \mathbf{S}^{j_{r+1}-1}\mathbf{x}_{r+1}$  together with the vectors in (i) are linearly independent. Then, since  $j_{r+1} \leq j_r$  by definition, if  $\mathbf{S}^{j_{r+1}}\mathbf{x}_{r+1} = \mathbf{0}$  we are through.

Suppose, then, that  $\mathbf{0} \neq \mathbf{S}^{j_{r+1}}\mathbf{x}_{r+1} = \sum \alpha_i \mathbf{S}^{v_i}\mathbf{x}_{\mu_i}$ , with all  $\alpha_i \neq 0$  and  $v_1 \leq v_2 \leq \dots$ . Clearly, we must have all  $\mu_i \leq r$ . For otherwise we could repeatedly multiply both sides by  $\mathbf{S}$ , substitute the above equality for  $\mathbf{S}^{j_{r+1}}\mathbf{x}_{r+1}$  when it appears on the right, and obtain all  $\mathbf{S}^{j_{k+1}+r}\mathbf{x}_{r+1} \neq \mathbf{0}$  for  $r = 0, 1, 2, \dots$  (since  $\alpha_i \neq 0$ ), contradicting  $\mathbf{S}^{k_1}\mathbf{x}_{r+1} = \mathbf{0}$ . If  $v_1 \leq j_{r+1} - 1$ , then  $j_{r+1} + \mu_1 - v_1 - 1 \geq \mu_1$ . But

$$\mathbf{S}^{j_{r+1}+\mu_1-v_1-1}\mathbf{x}_{r+1} = \alpha_1 \mathbf{S}^{\mu_1-1}\mathbf{x}_{\mu_1} + \dots$$

together with  $\mathbf{x}_{r+1}, \mathbf{S}\mathbf{x}_{r+1}, \dots, \mathbf{S}^{j_{r+1}+\mu_1-v_1-2}\mathbf{x}_{r+1}$ , and the first  $\mu_1 - 1$  rows of vectors in (i) are independent, contradicting (ii). Hence  $\mathbf{S}^{j_{r+1}}\mathbf{x}_{r+1}$  has the representation  $\mathbf{S}^{j_{r+1}}\mathbf{x}_{r+1} = \mathbf{S}^{j_{r+1}}(\sum \alpha_i \mathbf{S}^{v_i}\mathbf{x}_{\mu_i})$ , with all  $\mu_i \leq r$ . Hence the vector  $\mathbf{x}_{r+1} - \sum \alpha_i \mathbf{S}^{v_i}\mathbf{x}_{\mu_i}$  satisfies all the necessary requirements.

*The case of distinct eigenvalues.* If a matrix  $\mathbf{B}$  has  $n$  distinct eigenvalues, it is similar to a diagonal matrix  $\mathbf{A}$  given by its canonical form  $\mathbf{C}^{-1}\mathbf{B}\mathbf{C} = \mathbf{A}$ . In this case  $\mathbf{C}$  can be constructed directly from the eigenvectors of  $\mathbf{B}$  (which

form the basis  $\{\mathbf{y}^1, \dots, \mathbf{y}^n\}$  of the preceding section). For  $\mathbf{B}\mathbf{y}^i = \lambda_i \mathbf{y}^i$ , and if  $\mathbf{y}^i = \langle y_1^i, y_2^i, \dots, y_n^i \rangle$ , then, writing  $\mathbf{y}^i$  as a column vector, we have

$$\begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix} \begin{vmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_n^i \end{vmatrix} = \lambda_i \begin{vmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_n^i \end{vmatrix},$$

whence, by quick computation

$$\begin{aligned} & \begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix} \begin{vmatrix} y_1^1 & y_1^2 & \cdots & y_1^n \\ y_2^1 & y_2^2 & \cdots & y_2^n \\ \vdots & \vdots & & \vdots \\ y_n^1 & y_n^2 & \cdots & y_n^n \end{vmatrix} \\ &= \begin{vmatrix} y_1^1 & y_1^2 & \cdots & y_1^n \\ y_2^1 & y_2^2 & \cdots & y_2^n \\ \vdots & \vdots & & \vdots \\ y_n^1 & y_n^2 & \cdots & y_n^n \end{vmatrix} \begin{vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{vmatrix}. \end{aligned}$$

Since the matrix  $\mathbf{Y} = \|y_j^i\|$  maps the basic column vectors  $\langle 1, 0, \dots, 0 \rangle$ ,  $\langle 0, 1, 0, \dots, 0 \rangle, \dots$  onto  $\mathbf{y}^1, \mathbf{y}^2, \dots$ , respectively, it maps  $U^n$  onto itself; hence  $\mathbf{Y}^{-1}$  exists and  $\mathbf{Y}^{-1}\mathbf{B}\mathbf{Y} = \mathbf{\Lambda}$ , where  $\mathbf{\Lambda}$  denotes the above diagonal matrix of the eigenvalues.

### EXERCISES

1. Find the Jordan forms for the matrices

$$\begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & 1 \\ 2 & -1 & 0 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix}.$$

2. A common type of square matrix is one in which all elements are nonnegative and the sum of the elements in each row is unity. For such a matrix  $\mathbf{A}$  prove that 1 is an eigenvalue, that all eigenvalues are less than or equal to 1 in absolute value, and that the index of the eigenvalue 1 is 1; that is, the null space of  $\mathbf{A} - \mathbf{I}$  and  $(\mathbf{A} - \mathbf{I})^2$  coincide.

3. Construct an  $n \times n$  matrix  $\mathbf{A}$  with distinct eigenvalues such that  $\mathbf{A}^n = \mathbf{I}$ .

4. An equivalence relation on transformations is one which satisfies the following three postulates:

- (1) Any transformation is equivalent to itself.
- (2) If  $\mathbf{A}$  is equivalent to  $\mathbf{B}$ , then  $\mathbf{B}$  is equivalent to  $\mathbf{A}$ .
- (3) If  $\mathbf{A}$  is equivalent to  $\mathbf{B}$  and  $\mathbf{B}$  is equivalent to  $\mathbf{C}$ , then  $\mathbf{A}$  is equivalent to  $\mathbf{C}$ .

Prove that the notion of "similarity" is an equivalence relation.

5. In  $E^3$  with basis  $\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3$  let  $\mathbf{a}$  be a fixed vector with  $(\mathbf{a}, \mathbf{a}) = 1$ . Consider the hyperplane  $H = \{\mathbf{x} | (\mathbf{a}, \mathbf{x}) = 0\}$ . By the orthogonal projection  $\mathbf{y}'$  of a vector onto  $H$  we mean the unique vector  $\mathbf{y}'$  in  $H$  such that  $\mathbf{y}' = \mathbf{y} - \alpha\mathbf{a}$  for some scalar  $\alpha$ . Define  $\mathbf{P}\mathbf{y} = \mathbf{y}'$ . Then  $\mathbf{P}\mathbf{y} = \mathbf{y} - (\mathbf{y}, \mathbf{a})\mathbf{a} \cdot \mathcal{R}(P) = H$ . The null space of  $P$  is the space spanned by the vector  $\mathbf{a}$ . The eigenvalues of  $\mathbf{P}$  are 0 and 1, and the Jordan form of  $\mathbf{P}$  is given by

$$\begin{vmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

Verify all these statements.

**A.5 Transposed, normal, and hermitian matrices; orthogonal complement.** If  $\mathbf{A}$  is any complex  $m \times n$  matrix, the *conjugate transpose*  $\mathbf{A}^*$  of  $\mathbf{A}$  is the  $n \times m$  matrix with  $ij$ th entry

$$a_{ij}^* = \bar{a}_{ji} \quad (i = 1, \dots, n; j = 1, \dots, m).$$

Clearly  $(\mathbf{A} + \gamma\mathbf{B})^* = \mathbf{A}^* + \bar{\gamma}\mathbf{B}^*$ , and  $(\mathbf{AB}^*) = \mathbf{B}^*\mathbf{A}^*$  is easily verified when  $\mathbf{AB}$  makes sense. For  $\mathbf{x}$  in  $U^n$  and  $\mathbf{y}$  in  $U^m$

$$(\mathbf{Ax}, \mathbf{y}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j \bar{y}_i = \sum_{j=1}^n x_j \sum_{i=1}^m \bar{a}_{ij} y_i = (\mathbf{x}, \mathbf{A}^*\mathbf{y}).$$

As a consequence,  $\mathcal{R}(\mathbf{A})$  is orthogonal to  $\mathcal{N}(\mathbf{A}^*)$ ; that is,  $(\mathbf{Ax}, \mathbf{y}) = 0$  if  $\mathbf{A}^*\mathbf{y} = \mathbf{0}$ . Indeed,  $(\mathbf{Ax}, \mathbf{y}) = 0$  for all  $\mathbf{x}$  if and only if  $\mathbf{A}^*\mathbf{y} = \mathbf{0}$ . Hence, if we denote by  $\mathcal{R}(\mathbf{A})^\perp$  the set of all elements of  $U^m$  orthogonal to the subspace  $\mathcal{R}(\mathbf{A})$ , we may write  $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^*)$ ; similarly  $\mathcal{R}(\mathbf{A}^*)^\perp = \mathcal{N}(\mathbf{A})$ .  $\mathcal{R}(\mathbf{A})^\perp$  is known as the *orthogonal complement* of  $\mathcal{R}(\mathbf{A})$ . (The orthogonal complement of any subset of  $U^n$  is always a subspace; if  $V$  is a subspace, it is easily shown that  $V^{\perp\perp} = (V^\perp)^\perp = V$ .) When we are dealing with the real vector space  $E^n$ , the conjugate transpose matrix  $\mathbf{A}^*$  reduces to the ordinary transpose matrix  $\mathbf{A}'$  (obtained from  $\mathbf{A}$  by interchanging corresponding rows and columns), and all the corresponding operations and relations of  $\mathbf{A}'$ ,  $\mathcal{R}(\mathbf{A}')^\perp$ , and  $\mathcal{N}(\mathbf{A}')$  remain valid.

When  $\mathbf{A}$  is  $n \times n$ ,  $\mathbf{A}$  is called *normal* if  $\mathbf{A}^*\mathbf{A} = \mathbf{AA}^*$  (in general,  $\mathbf{A}^*\mathbf{A}$  is  $n \times n$  and  $\mathbf{AA}^*$  is  $m \times m$ ). One form of normal matrix is the hermitian

matrix:  $\mathbf{A}$  is *hermitian* if  $\mathbf{A} = \mathbf{A}^*$ . (A real hermitian matrix is called symmetric; here  $\mathbf{A} = \mathbf{A}'$ .) Since  $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$ , and clearly  $\mathbf{A}^{**} = \mathbf{A}$ , each  $n \times n$  matrix  $\mathbf{A}$  has  $\mathbf{A}^*\mathbf{A}$  (or  $\mathbf{AA}^*$ ) hermitian.

Normal matrices have elegant properties compared with general matrices. As we shall see, the eigenvectors of a normal matrix  $\mathbf{A}$  span  $U^n$  and those corresponding to distinct eigenvalues are orthogonal, so that we may obtain an orthonormal basis of eigenvectors for  $U^n$ ; moreover,  $\mathbf{Ax} = \lambda\mathbf{x}$  ( $\mathbf{x} \neq \mathbf{0}$ ) is equivalent to  $\mathbf{A}^*\mathbf{x} = \bar{\lambda}\mathbf{x}$ , that is,  $\mathbf{A}$  and  $\mathbf{A}^*$  share the same eigenvectors with conjugate eigenvalues.

To obtain these facts, we begin by noting that when  $\mathbf{A}$  is normal,  $\mathbf{B} = \mathbf{A} - \lambda\mathbf{I}$  is also normal. Thus if  $\mathbf{B}^k\mathbf{x} = \mathbf{0}$ , we have  $\mathbf{B}^{*k}\mathbf{B}^k = \mathbf{0}$ ; and since  $\mathbf{B}^*$  and  $\mathbf{B}$  commute,  $(\mathbf{B}^*\mathbf{B})^k\mathbf{x} = \mathbf{0}$ , and surely  $(\mathbf{B}^*\mathbf{B})^{2^k}\mathbf{x} = \mathbf{0}$ . Let  $\mathbf{C} = \mathbf{B}^*\mathbf{B} = \mathbf{C}^*$ . Then

$$(\mathbf{C}^{2^k}\mathbf{x}, \mathbf{x}) = 0 = (\mathbf{C}^{2^{k-1}}\mathbf{x}, \mathbf{C}^{2^{k-1}}\mathbf{x}),$$

and thus  $\mathbf{C}^{2^{k-1}}\mathbf{x} = \mathbf{0}$ ; similarly  $\mathbf{C}^{2^{k-2}}\mathbf{x} = \mathbf{0}$ , and continuing, we find ultimately that  $\mathbf{Cx} = \mathbf{0}$ , or  $\mathbf{B}^*\mathbf{Bx} = \mathbf{0}$ . But again  $(\mathbf{B}^*\mathbf{Bx}, \mathbf{x}) = 0 = (\mathbf{Bx}, \mathbf{Bx})$ , and thus  $\mathbf{Bx} = \mathbf{0}$ . Consequently each element of  $\mathbf{x}_i$  obtained in our Jordan decomposition is an eigenvector of  $\mathbf{A}$ , and  $U^n$  is spanned by eigenvectors. Further, if  $\mathbf{Bx} = \mathbf{0}$ , then

$$(\mathbf{Bx}, \mathbf{Bx}) = 0 = (\mathbf{B}^*\mathbf{Bx}, \mathbf{x}) = (\mathbf{B}\mathbf{B}^*\mathbf{x}, \mathbf{x}) = (\mathbf{B}^*\mathbf{x}, \mathbf{B}^*\mathbf{x}),$$

and thus  $\mathbf{B}^*\mathbf{x} = \mathbf{0}$ . In terms of  $\mathbf{A} - \lambda\mathbf{I}$ , if  $\mathbf{Ax} = \lambda\mathbf{x}$ , then  $\mathbf{A}^*\mathbf{x} = \bar{\lambda}\mathbf{x}$ ; applying this to  $\{\mathbf{A}^*, \bar{\lambda}\}$  in place of  $\{\mathbf{A}, \lambda\}$  yields the converse.

Consequently, if  $\mathbf{Ax} = \lambda\mathbf{x}$  and  $\mathbf{Ay} = \mu\mathbf{y}$ , with  $\lambda \neq \mu$  and  $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$ , we have

$$(\mathbf{Ax}, \mathbf{y}) = (\lambda\mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{A}^*\mathbf{y}) = (\mathbf{x}, \bar{\mu}\mathbf{y}) = \mu(\mathbf{x}, \mathbf{y}).$$

It follows that  $(\lambda - \mu)(\mathbf{x}, \mathbf{y}) = 0$ ; thus  $(\mathbf{x}, \mathbf{y}) = 0$ , and distinct eigenvalues have orthogonal eigenvectors. Since the set of eigenvectors corresponding to  $\lambda_i$  forms the subspace  $Y_i$  when  $\mathbf{0}$  is adjoined, we can find an orthonormal basis for each  $Y_i$ ; obviously the union of these bases is the required orthonormal basis for  $U^n$ .

Applying the construction of the last section, we find that  $\mathbf{Y}^{-1}\mathbf{AY} = \mathbf{A}$ , a diagonal matrix, where the column vectors of  $\mathbf{Y}$  are orthonormal (being our basis of orthonormal eigenvectors). Since this amounts to saying

$$\sum_{j=1}^n y_j^i \bar{y}_j^k = \delta_{ik},$$

we have  $\mathbf{YY}^* = \mathbf{I}$ , and thus  $\mathbf{Y}^{-1} = \mathbf{Y}^*$ . Such a matrix is called *unitary*,

and  $\mathbf{A}$  is said to be unitarily equivalent (rather than unitarily similar) to  $\mathbf{A}$ . A unitary matrix preserves inner products (hence length and orthogonality), for  $(\mathbf{Y}\mathbf{x}, \mathbf{Y}\mathbf{y}) = (\mathbf{Y}^*\mathbf{Y}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})$ .

Hermitian matrices have the further property that each  $\lambda_i$  is real, for the equal matrices  $\mathbf{A}$  and  $\mathbf{A}^*$  have conjugate eigenvalues; moreover, any normal matrix with this property is hermitian, for then

$$\mathbf{A} = \mathbf{Y}\mathbf{\Lambda}\mathbf{Y}^* = (\mathbf{Y}\mathbf{\Lambda}^*\mathbf{Y}^*)^* = (\mathbf{Y}\mathbf{\Lambda}\mathbf{Y}^*)^* = \mathbf{A}^*.$$

A unitary matrix is of course normal, and each eigenvalue is of unit modulus ( $|\lambda| = 1$ ; for  $\mathbf{U}\mathbf{x} = \lambda\mathbf{x}$  has the same length as  $\mathbf{x}$ ). Conversely, any normal matrix with unimodular eigenvalues is unitary. A real unitary matrix is called *orthogonal*; any real eigenvalue is  $\pm 1$  and has corresponding real eigenvectors, which either are fixed points of the corresponding transformation or are reflected into their negatives. (No real eigenvalues need exist if  $n$  is even.) An orthogonal matrix also preserves angles.

*Projections.* If  $U^n = V \oplus W$ , then the unique expression of  $\mathbf{x}$  as  $\mathbf{v} + \mathbf{w}$  ( $\mathbf{v} \in V$ ;  $\mathbf{w} \in W$ ) leads to a linear transformation  $\mathbf{P}$  mapping  $\mathbf{x}$  into  $\mathbf{v}$ , and  $\mathbf{P}^2\mathbf{x} = \mathbf{P}\mathbf{v} = \mathbf{v} = \mathbf{P}\mathbf{x}$ , that is,  $\mathbf{P}^2 = \mathbf{P}$ . If  $V$  and  $W$  are orthogonal subspaces ( $W = V^\perp$ ), then  $\mathbf{P} = \mathbf{P}^2 = \mathbf{P}^*$ . Conversely, any transformation  $\mathbf{P}$  such that  $\mathbf{P} = \mathbf{P}^2 = \mathbf{P}^*$  (called an *orthogonal projection*, or *projection*) yields a direct-sum decomposition  $U^n = \mathcal{R}(\mathbf{P}) + \mathcal{N}(\mathbf{P})$ . As a normal matrix,  $\mathbf{P}$  can be represented (by an appropriate choice of orthonormal basis) as a diagonal matrix with 0's and 1's on the main diagonal and 0's elsewhere.

### EXERCISES

1. Interpret and prove the statement that any orthogonal transformation may be regarded as a rotation of axes about a fixed origin.
2. We have defined  $\mathbf{A}^*$  with respect to a particular basis of  $U^n$ , but it may also be defined directly from the inner product. Prove that for each  $\mathbf{y} \in U^n$ ,  $\mathbf{A}^*\mathbf{y}$  is the unique vector  $\mathbf{y}'$  such that  $(\mathbf{A}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y}')$  for all  $\mathbf{x} \in U^n$ .
3. On the space  $\mathcal{P}^n$  of polynomials with real coefficients over the interval  $[a, b]$ , let  $f(t)$  be a specific even polynomial. Define the linear transformation  $\mathbf{A}$  on  $\mathcal{P}^n$  as follows:

$$(\mathbf{A}p)(t) = \int_a^b f(t - y)p(y) dy.$$

Show that  $\mathbf{A}$  is a symmetric transformation.

4. Prove that a matrix is unitary if and only if it transforms orthonormal vectors into orthonormal vectors.
5. Let  $\mathbf{A}$  be a normal transformation with eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_r$  (eigenvalues of multiplicity higher than one appearing only once). Let  $\mathbf{E}_k$  be the per-

pendicular projection on the subspace of eigenvectors of  $\lambda_k$ . Show that if  $i \neq j$ , then  $\mathbf{E}_i \mathbf{E}_j = 0$ . Show also that  $\mathbf{E}_1 + \mathbf{E}_2 + \cdots + \mathbf{E}_k$  is a perpendicular projection for all  $k \leq r$ , and that  $\mathbf{E}_1 + \mathbf{E}_2 + \cdots + \mathbf{E}_r = \mathbf{I}$ .

6. For the same setup as in Problem 5, prove the matrix relation

$$\mathbf{A} = \lambda_1 \mathbf{E}_1 + \lambda_2 \mathbf{E}_2 + \cdots + \lambda_r \mathbf{E}_r.$$

**A.6 Quadratic form.** Any expression

$$\sum_{i,j=1}^n a_{ij} x_i x_j$$

in the  $n$  real variables  $x_1, \dots, x_n$  (where  $a_{ij}$  is a real constant) is called a quadratic form. Since

$$Q(\mathbf{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j = \sum \frac{1}{2} (a_{ij} + a_{ji}) x_i x_j,$$

one can replace the original  $a_{ij}$  by a symmetric array with  $a_{ij} = a_{ji}$ . The corresponding matrix  $\mathbf{A}$  is then a hermitian transformation of  $U^n$  into itself and has only real eigenvalues. Let  $\lambda$  denote one of these eigenvalues. Then since  $\mathbf{A}(\mathbf{x} + i\mathbf{y}) = \lambda(\mathbf{x} + i\mathbf{y})$  implies  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  and  $\mathbf{A}\mathbf{y} = \lambda\mathbf{y}$ , and one of these real vectors is nonzero,  $\lambda$  has a corresponding eigenvector in  $E^n$ . Repeating our former arguments (applying them now to  $E^n$  instead of  $U^n$ ), we obtain a real orthonormal basis  $\{\mathbf{y}^1, \dots, \mathbf{y}^n\}$  of eigenvectors corresponding to the eigenvalues  $\lambda_1, \dots, \lambda_n$ .

Further, the matrix  $\mathbf{Y}$  we obtain from  $\mathbf{y}^1, \dots, \mathbf{y}^n$  is now a real unitary (i.e., orthogonal) matrix, which yields  $\mathbf{Y}^* \mathbf{A} \mathbf{Y} = \mathbf{\Lambda}$  (diagonal), or  $\mathbf{A} = \mathbf{Y} \mathbf{\Lambda} \mathbf{Y}^*$ . Consequently, if we write

$$\mathbf{x} = \sum_{i=1}^n x'_i \mathbf{y}^i \quad (\mathbf{x} \in E^n),$$

from which it follows that

$$\mathbf{Y}\mathbf{x}' = \sum_{i=1}^n x'_i \mathbf{y}^i = \mathbf{x},$$

then  $\mathbf{Y}^* \mathbf{x} = \mathbf{Y}^{-1} \mathbf{x} = \mathbf{x}'$  and we have

$$Q(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x}) = (\mathbf{Y} \mathbf{\Lambda} \mathbf{Y}^* \mathbf{x}, \mathbf{x}) = (\mathbf{\Lambda} \mathbf{Y}^* \mathbf{x}, \mathbf{Y}^* \mathbf{x}) = \sum_{i=1}^n \lambda_i (x'_i)^2.$$

Thus, by selecting an appropriate orthonormal basis we may express  $Q$  in diagonal form, with no mixed products of components appearing. Geometrically, then, the locus of all  $\mathbf{x}$  satisfying  $Q(\mathbf{x}) = 1$  is an ellipsoid if all

$\lambda_i > 0$ . This occurs if and only if  $Q(\mathbf{x})$  is nonnegative and  $Q(\mathbf{x}) = 0$  implies  $\mathbf{x} = \mathbf{0}$ , i.e., when  $Q$  is *positive definite*.

Two quadratic forms  $Q$  and  $Q_1$  are said to be unitarily equivalent if one arises from the other via a unitary transformation of  $E^n$  into itself, i.e., if  $Q(\mathbf{x}) = Q_1(\mathbf{Ux})$ . (In our discussion  $\mathbf{U}$  is of course an orthogonal matrix, but similar results are available for complex quadratic forms.) This occurs if and only if  $\mathbf{A}$  and  $\mathbf{A}_1$  are unitarily equivalent, for then  $(\mathbf{Ax}, \mathbf{x}) = (\mathbf{A}_1 \mathbf{Ux}, \mathbf{Ux}) = (\mathbf{U}^* \mathbf{A}_1 \mathbf{Ux}, \mathbf{x})$  and it follows that  $\mathbf{A} = \mathbf{U}^* \mathbf{A}_1 \mathbf{U}$ . Thus  $Q$  and  $Q_1$  are unitarily equivalent if and only if their diagonal forms feature the same eigenvalues (of identical multiplicities, but possibly in a different arrangement).

If  $Q$  is a positive definite quadratic form, then one can replace the usual inner product by

$$[\mathbf{x}, \mathbf{y}] = (\mathbf{Ax}, \mathbf{y}),$$

where  $\mathbf{A}$  is the corresponding hermitian matrix; for  $[\mathbf{x}, \mathbf{y}]$  will have all the required properties. Geometrically, this amounts to replacing the unit sphere  $(\mathbf{x}, \mathbf{x}) = 1$  by an ellipsoid. In terms of the components  $x'_1, x'_2, \dots, x'_n$  of  $\mathbf{x}$  relative to this basis, the new length of  $\mathbf{x}$  is computed as the ordinary length of  $(\sqrt{\lambda_1} x'_1, \dots, \sqrt{\lambda_n} x'_n)$ .

The new inner product will of course give rise to a new version of the conjugate transpose of a matrix  $\mathbf{A}$  if one insists that  $[\mathbf{Ax}, \mathbf{y}] = [\mathbf{x}, \mathbf{A}^* \mathbf{y}]$ . Thus one obtains an analogous concept of normal and hermitian linear transformations relative to the new inner product; these are of course related to the old by virtue of the obvious change of scale (in each  $x'_i$ ) which transforms our new inner product into the old.

### EXERCISES

1. Show that for any real matrix transformation  $\mathbf{A}$  on  $E^n$ ,

$$2(\mathbf{Ax}, \mathbf{y}) = (\mathbf{A}(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) - (\mathbf{Ax}, \mathbf{x}) - (\mathbf{Ay}, \mathbf{y}).$$

With the aid of this identity prove that if  $(\mathbf{Ax}, \mathbf{x}) = (\mathbf{Bx}, \mathbf{x})$  for all  $\mathbf{x}$  in  $E^n$ , then  $\mathbf{A} = \mathbf{B}$ .

2. Prove that  $[\mathbf{x}, \mathbf{y}]$  as defined in this section satisfies the requirements for an inner product. Is it necessary that  $\mathbf{A}$  be hermitian?

3. Prove that if

$$\sum_{i=1}^n x_i^2 = Q_1 + Q_2,$$

where  $Q_1$  and  $Q_2$  are quadratic forms of  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  of ranks  $r$  and  $n - r$ , respectively, then there is an orthogonal transformation  $\mathbf{C}$  such that

$$\mathbf{x} = \mathbf{Cy}, \quad Q_1 = \sum_{i=1}^r y_i^2, \quad Q_2 = \sum_{i=r+1}^n y_i^2.$$

4. Prove that if  $Q$  is positive definite, then the locus of all  $\mathbf{x}$  such that  $Q(\mathbf{x}) = \text{constant}$  is an ellipsoid with semi-axes proportional to  $1/\sqrt{\lambda_r}$ , where the  $\lambda_r$ 's are the eigenvalues of  $A$ . Any direction corresponding to an eigenvector of  $\lambda_r$  is a semi-axis of the ellipsoid.

**A.7 Matrix-valued functions.** For any  $n \times n$  matrix  $\mathbf{A}$  and polynomial  $p(t) = a_0 + a_1t + \cdots + a_k t^k$  one may clearly form  $a_0\mathbf{I} + a_1\mathbf{A} + \cdots + a_k\mathbf{A}^k = p(\mathbf{A})$ ; thus we obtain a matrix-valued function of  $\mathbf{A}$ ; that is,  $\mathbf{A}$  maps into  $p(\mathbf{A})$ . Similarly, for appropriate power series

$$f(t) = \sum_{i=0}^{\infty} a_i t^i$$

the corresponding matrix series

$$f(\mathbf{A}) = \sum_{i=0}^{\infty} a_i \mathbf{A}^i$$

(where  $\mathbf{A}^0 = \mathbf{I}$ ) will be convergent [in the sense that each entry of the partial sums converges, with its limit the value of the corresponding entry in  $f(\mathbf{A})$ ]. For example, if  $f$  is an entire function,  $f(\mathbf{A})$  will be defined for every square matrix; in particular,

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{1}{2!} \mathbf{A}^2 + \cdots$$

is well defined. If  $\lambda_1, \dots, \lambda_k$  are the eigenvalues of  $\mathbf{A}$ , then  $f(\lambda_1), \dots, f(\lambda_k)$  are the eigenvalues of  $f(\mathbf{A})$ . For  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  implies  $\mathbf{A}^2\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x}$ ,  $\mathbf{A}^3\mathbf{x} = \lambda^3\mathbf{x}$ ,  $\dots$ ; hence  $f(\mathbf{A})\mathbf{x} = f(\lambda)\mathbf{x}$ , so that each  $f(\lambda_i)$  is an eigenvalue of  $f(\mathbf{A})$  whose associated eigenvectors are those belonging to the eigenvalue  $\lambda_i$  of  $\mathbf{A}$ . To see that all eigenvalues of  $f(\mathbf{A})$  are so obtained, consider the Jordan canonical form  $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$  of  $\mathbf{A}$ , noting that  $\mathbf{B}^{-1}\mathbf{A}^k\mathbf{B} = (\mathbf{B}^{-1}\mathbf{A}\mathbf{B})^k$  and thus  $\mathbf{B}^{-1}f(\mathbf{A})\mathbf{B} = f(\mathbf{B}^{-1}\mathbf{A}\mathbf{B})$ ; the computation of  $f(\mathbf{B}^{-1}\mathbf{A}\mathbf{B})$  can be made sub-block by sub-block, and its eigenvalues become apparent.

A particularly important example of a matrix-valued function is furnished by  $(\lambda\mathbf{I} - \mathbf{A})^{-1}$ . When  $|\lambda|$  exceeds the modulus of each eigenvalue, the formal power series expansion

$$\begin{aligned} (\lambda\mathbf{I} - \mathbf{A})^{-1} &= \lambda^{-1} \left( \mathbf{I} - \frac{1}{\lambda} \mathbf{A} \right)^{-1} \\ &= \lambda^{-1} \left( \mathbf{I} + \frac{1}{\lambda} \mathbf{A} + \frac{1}{\lambda^2} \mathbf{A}^2 + \cdots \right) = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{\lambda^{i+1}} \end{aligned}$$

is valid. Moreover, by fixing  $\mathbf{A}$  we obtain an analytic function of the complex variable  $\lambda$ ; that is,  $\lambda$  maps into  $(\lambda\mathbf{I} - \mathbf{A})^{-1}$ , defined for all  $\lambda \neq \lambda_1, \dots, \lambda_k$ . The function is analytic in the sense that

$$\begin{aligned}\frac{d}{d\lambda}(\lambda\mathbf{I} - \mathbf{A})^{-1} &= \lim_{\lambda' \rightarrow \lambda} \frac{1}{\lambda' - \lambda} [(\lambda'\mathbf{I} - \mathbf{A})^{-1} - (\lambda\mathbf{I} - \mathbf{A})^{-1}] \\ &= -(\lambda\mathbf{I} - \mathbf{A})^{-2}\end{aligned}$$

exists where  $\lambda \neq \lambda_1, \dots, \lambda_k$ . One may also apply the Cauchy integral theorem, its validity being obtained in the usual fashion; here, however, we can interpret the function's analyticity as meaning that it has a power series expansion about each  $\lambda \neq \lambda_1, \dots, \lambda_k$ , with matrix coefficients. Each  $\lambda_i$  is a pole of the function; for if  $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$  is again the Jordan canonical form, we have  $\mathbf{A} = \mathbf{B}\mathbf{C}\mathbf{B}^{-1}$  and

$$\lambda\mathbf{I} - \mathbf{A} = \lambda\mathbf{I} - \mathbf{B}\mathbf{C}\mathbf{B}^{-1} = \mathbf{B}(\lambda\mathbf{I} - \mathbf{C})\mathbf{B}^{-1};$$

hence  $(\lambda\mathbf{I} - \mathbf{A})^{-1} = \mathbf{B}(\lambda\mathbf{I} - \mathbf{C})^{-1}\mathbf{B}^{-1}$ , and it suffices to consider  $(\lambda\mathbf{I} - \mathbf{C})^{-1}$ , or indeed  $(\mathbf{C} - \lambda\mathbf{I})^{-1}$ . Since  $\mathbf{C} - \lambda\mathbf{I}$  has the same block-diagonal form as  $\mathbf{C}$ , and since block-diagonal matrices (with blocks of similar size) multiply block by block—i.e., since

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{AC} & \mathbf{0} \\ \mathbf{0} & \mathbf{BD} \end{pmatrix}$$

if  $\mathbf{A}$  and  $\mathbf{C}$  are  $k \times k$ , and  $\mathbf{B}$  and  $\mathbf{D}$  are  $l \times l$ —it is sufficient to compute the inverse of each sub-block

$$\begin{vmatrix} \lambda_i - \lambda & 0 & \cdots & \cdots & 0 \\ 1 & \lambda_i - \lambda & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & \cdots & \lambda_i - \lambda \end{vmatrix}$$

of  $\mathbf{C} - \lambda\mathbf{I}$ . Writing

$$\mathbf{D} = (\lambda_i - \lambda)\mathbf{I} + \mathbf{E} = (\lambda_i - \lambda) \left( \mathbf{I} + \frac{1}{\lambda_i - \lambda} \mathbf{E} \right),$$

where  $\mathbf{E}$  is the  $k \times k$  matrix with 1's below the main diagonal and 0's elsewhere, and formally expanding in analogy with  $(1 + x)^{-1} = 1 - x + x^2 - x^3 + \dots$ , we obtain

$$\begin{aligned}\mathbf{D}^{-1} &= (\lambda_i - \lambda)^{-1} \left( \mathbf{I} + \frac{1}{\lambda_i - \lambda} \mathbf{E} \right)^{-1} \\ &= (\lambda_i - \lambda)^{-1} \left( \mathbf{I} - \frac{1}{\lambda_i - \lambda} \mathbf{E} + \frac{1}{(\lambda_i - \lambda)^2} \mathbf{E}^2 - \dots \right),\end{aligned}$$

which converges, since  $\mathbf{E}^k = \mathbf{0}$ . Moreover, for the same reason we may check that the value is the appropriate inverse by multiplying by

$$\mathbf{I} + \frac{1}{\lambda_i - \lambda} \mathbf{E}.$$

Thus the highest power of  $(\lambda_i - \lambda)^{-1}$  which appears is the  $k$ th, and  $\mathbf{D}^{-1}$  appears as a Laurent expansion (with constant matrix coefficients) corresponding to a pole of order  $k$  at  $\lambda_i$ . Consequently the order of the pole we obtain at  $\lambda_i$  for  $(\lambda\mathbf{I} - \mathbf{C})^{-1}$  [or  $(\lambda\mathbf{I} - \mathbf{A})^{-1}$ ] is the size of the largest sub-block featuring  $\lambda_i$ ; recall that this is the least  $k$  for which

$$\mathfrak{N}[(\lambda\mathbf{I} - \mathbf{A})^k] = \mathfrak{N}[(\lambda\mathbf{I} - \mathbf{A})^{k+1}].$$

### EXERCISES

1. Show that if  $\mathbf{AB} = \mathbf{BA}$ , then  $e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}}e^{\mathbf{B}}$ .
2. Find a Laurent series expansion for  $(\lambda\mathbf{I} - \mathbf{A})^{-n}$  and show that it converges when  $\lambda$  exceeds the modulus of each eigenvalue. Under these conditions, show that

$$\frac{d}{d\lambda} (\lambda\mathbf{I} - \mathbf{A})^{-n} = -n(\lambda\mathbf{I} - \mathbf{A})^{-n-1}.$$

3. Let  $\lambda_1, \lambda_2, \dots, \lambda_r$  and  $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r$  (as defined in Problem 5 of Section A.5) be the eigenvalues and perpendicular projections corresponding to a normal matrix  $\mathbf{A}$ . Let  $f(\lambda)$  be a convergent power series. Show that

$$f(\mathbf{A}) = \sum_{j=1}^r f(\lambda_j) \mathbf{E}_j.$$

4. Show that if  $f(d) = \mathbf{A}_0 + \mathbf{A}_1d + \mathbf{A}_2d^2 + \dots$  converges absolutely in some interval about  $d_0$ , then  $f'(d_0) = \mathbf{A}_1 + 2\mathbf{A}_2d_0 + 3\mathbf{A}_3d_0^2 + \dots$ .
5. If  $\mathbf{A}$  is a normal matrix with nonzero eigenvalues, how can you define  $\log \mathbf{A}$ ?

**A.8 Determinants; minors, cofactors.** The determinant  $|\mathbf{A}|$  (sometimes written  $\det \mathbf{A}$ ) of an  $n \times n$  matrix  $\mathbf{A}$  is given by the sum

$$\sum_{\pi} \delta(\pi) a_{1,\pi(1)} a_{2,\pi(2)} \dots a_{n,\pi(n)} = \sum_{\pi} \delta(\pi) a_{\pi(1),1} a_{\pi(2),2} \dots a_{\pi(n),n}, \quad (\text{A.8.1})$$

where  $\pi$  runs over all permutations of  $\{1, 2, \dots, n\}$ , and  $\delta(\pi)$  is 1 if  $\pi$  is an even permutation and  $-1$  if  $\pi$  is an odd permutation. Interchange of any two rows (or columns) of  $\mathbf{A}$  produces a matrix  $\mathbf{B}$  such that  $|\mathbf{B}| = -|\mathbf{A}|$ , since the corresponding terms in (A.8.1) occur with permutations of opposite parity; thus if  $\mathbf{A}$  has two identical rows (or columns),  $|\mathbf{A}| = 0$ . The

determinant of the matrix obtained by deleting the  $i$ th row and  $j$ th column of  $\mathbf{A}$  is the *minor*  $m_{ij}$  of  $a_{ij}$ ; the *cofactor*  $c_{ij}$  of  $a_{ij}$  is then defined by  $c_{ij} = (-1)^{i+j} m_{ij}$ .

It follows directly from (A.8.1) that

$$|\mathbf{A}| = \sum_{k=1}^n a_{ik} c_{ik} = \sum_{k=1}^n a_{kj} c_{kj} \quad (\text{A.8.2})$$

for all  $i, j$ ; these are the so-called row and column expansions of  $|\mathbf{A}|$ . Further, if  $i \neq j$ ,

$$\sum_{k=1}^n a_{ik} c_{jk} = 0 = \sum_{k=1}^n a_{kj} c_{ik},$$

since these expressions are the row and column expansions of  $|\mathbf{B}|$  where  $\mathbf{B}$  is a matrix with two identical rows or columns. Consequently

$$|\mathbf{A}| = \sum_{k=1}^n a_{ik} c_{ik} - \lambda \sum_{k=1}^n a_{jk} c_{ik} = \sum_{k=1}^n (a_{ik} - \lambda a_{jk}) c_{ik}$$

if  $i \neq j$ , so that  $|\mathbf{A}|$  coincides with the value of the determinant of the matrix obtained from  $\mathbf{A}$  by subtracting  $\lambda$  times the  $j$ th row from the  $i$ th row. Writing  $|\mathbf{A}|$  in the descriptive form

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

we may thus subtract any multiple of a row (or column) of the determinant from any other without changing its value.

Let

$$\mathbf{A} \begin{pmatrix} i_1, i_2, \dots, i_r \\ j_1, j_2, \dots, j_r \end{pmatrix}$$

denote the determinant of the matrix obtained from  $\mathbf{A}$  by deleting all rows but  $i_1, i_2, \dots, i_r$  and all columns but  $j_1, \dots, j_r$ , where  $i_1 < i_2 < \dots < i_r, j_1 < j_2 < \dots < j_r$ ; that is,

$$\mathbf{A} \begin{pmatrix} i_1, i_2, \dots, i_r \\ j_1, j_2, \dots, j_r \end{pmatrix} = \begin{vmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \dots & a_{i_1 j_r} \\ a_{i_2 j_1} & \dots & \dots & \dots \\ \vdots & & & \\ a_{i_r j_1} & \dots & \dots & a_{i_r j_r} \end{vmatrix}.$$

Such subdeterminants of a product  $\mathbf{AB}$  are related to those of  $\mathbf{A}$  and  $\mathbf{B}$  by the expansion

$$\begin{aligned} \mathbf{AB} & \left( \begin{matrix} i_1, i_2, \dots, i_r \\ j_1, j_2, \dots, j_r \end{matrix} \right) \\ &= \sum_{k_1 < k_2 < \dots < k_r} \mathbf{A} \left( \begin{matrix} i_1, i_2, \dots, i_r \\ k_1, k_2, \dots, k_r \end{matrix} \right) \mathbf{B} \left( \begin{matrix} k_1, k_2, \dots, k_r \\ j_1, j_2, \dots, j_r \end{matrix} \right), \quad (\text{A.8.3}) \end{aligned}$$

where the sum extends over all  $\{k_1, k_2, \dots, k_r\}$  satisfying the indicated relation.

From (A.8.2) and our succeeding remarks, if  $\text{adj}(\mathbf{A})$  denotes the matrix with  $ij$ th entry

$$b_{ij} = c_{ji},$$

the *adjoint matrix* of  $\mathbf{A}$ , then  $\mathbf{A} \cdot \text{adj}(\mathbf{A}) = |\mathbf{A}| \mathbf{I} = \text{adj}(\mathbf{A}) \cdot \mathbf{A}$ . Thus if  $|\mathbf{A}| \neq 0$ ,

$$\frac{1}{|\mathbf{A}|} \text{adj}(\mathbf{A})$$

is the inverse of  $\mathbf{A}$ . But  $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$  by (A.8.3); thus if  $\mathbf{A}^{-1}$  exists,  $|\mathbf{A}||\mathbf{A}^{-1}| = |\mathbf{I}| = 1$  and  $|\mathbf{A}| \neq 0$ . *Hence the nonsingular matrices are those with nonzero determinants.*

If  $\mathbf{B}^{-1}\mathbf{AB}$  is the Jordan canonical form of  $\mathbf{A}$ , then, since  $|\mathbf{B}^{-1}\mathbf{AB}| = |\mathbf{B}^{-1}||\mathbf{A}||\mathbf{B}| = |\mathbf{A}|$ , it follows that  $|\mathbf{A}|$  is the product of the powers of its eigenvalues. Consequently, if 0 appears among the eigenvalues (say  $\lambda_i = 0$ ), we may slightly change  $\lambda_i$  so as to have nonzero eigenvalues and a nonsingular matrix; the corresponding changes in the entries of  $\mathbf{A}$  are small, and thus *each singular  $\mathbf{A}$  is a limit of nonsingular matrices*.

Further, since

$$|\mathbf{B}^{-1}\mathbf{AB} - \lambda\mathbf{I}| = |\mathbf{B}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{B}| = |\mathbf{A} - \lambda\mathbf{I}|,$$

with  $\mathbf{B}$  as above, it follows that

$$|\mathbf{B}^{-1}\mathbf{AB} - \lambda\mathbf{I}| = \prod_{i=1}^k (\lambda_i - \lambda)^{n_i}.$$

Thus the  $\lambda_i$  are exactly the roots of the polynomial equation

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

in  $\lambda$ , the *characteristic equation* of  $\mathbf{A}$ . The multiplicity of  $\lambda_i$  as a root is called the *algebraic multiplicity* of  $\lambda_i$  (the algebraic multiplicity  $\geq$  the

geometric multiplicity = number of linearly independent eigenvectors corresponding to  $\lambda_i$ ; it is of course the number of times  $\lambda_i$  appears on the main diagonal of the Jordan form.

### EXERCISES

1. Show that if  $\langle x_1, y_1 \rangle$  and  $\langle x_2, y_2 \rangle$  are distinct points on the real plane, the equation

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0$$

represents a straight line through the points.

2. Prove the following statement:

Given a nonsingular  $n \times n$  matrix  $\mathbf{A}$  and a vector  $\mathbf{y}$  in  $U^n$ , let  $\mathbf{A}_{iy}$  be the matrix in which column  $i$  has been replaced by  $\mathbf{y}$  and all other columns remain unchanged. Then the solution of the system of equations  $\mathbf{Ax} = \mathbf{y}$ , where  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ , is given by

$$x_i = \frac{|\mathbf{A}_{iy}|}{|\mathbf{A}|} \quad (i = 1, \dots, n).$$

3. Show that if all the elements in a matrix above the main diagonal are zero, the diagonal elements are the eigenvalues.

4. Let  $\mathbf{A}$  have eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$ , with repetitions if necessary. Find  $|(I - \lambda\mathbf{A})^{-1}|$ .

5. By  $|\mathbf{x}|$  we mean  $(\mathbf{x}, \mathbf{x})^{\frac{1}{2}}$ . We say that a matrix or transformation  $\mathbf{A}$  is bounded on  $U$  if there is a constant  $K$  such that  $|\mathbf{Ax}| \leq K|\mathbf{x}|$  for all  $\mathbf{x}$  in  $U$ . The greatest lower bound,  $K_0$ , of all such  $K$ 's is the bound of  $\mathbf{A}$ . Prove that all transformations on  $U^n$  are bounded. Show that if  $\mathbf{A}$  is the limit of nonsingular matrices, then  $\mathbf{A}$  is singular if and only if the bounds of the inverses of these matrices approach infinity.

**A.9 Some identities.** Let  $\mathbf{U}$  be the  $n \times n$  matrix with all entries 1, and let  $\mathbf{u} = \langle 1, 1, \dots, 1 \rangle$  be an  $n$ -component vector of  $\mathbf{U}$  with all entries 1. The following identities relating the determinants and adjoints of  $\mathbf{A}$  and  $\mathbf{A} + a\mathbf{U}$  are used in Section 2.4:

$$\begin{aligned} |\mathbf{A} + a\mathbf{U}| &= |\mathbf{A}| + a(\mathbf{u}, \text{adj}(\mathbf{A})\mathbf{u}), \\ \mathbf{u} \text{adj}(\mathbf{A} + a\mathbf{U}) &= \mathbf{u} \text{adj}(\mathbf{A}), \\ \text{adj}(\mathbf{A} + a\mathbf{U})\mathbf{u} &= \text{adj}(\mathbf{A})\mathbf{u}. \end{aligned}$$

Here  $(\mathbf{u}, \text{adj}(\mathbf{A})\mathbf{u})$  when calculated yields

$$\sum_{i,j=1}^n A_{ij},$$

where  $A_{ij}$  is the cofactor of  $a_{ij}$ . If we regard  $|\mathbf{A}|$  as a function of the  $n^2$  variables  $a_{ij}$ , the row expansion

$$|\mathbf{A}| = \sum_{k=1}^n a_{ik} A_{ik}$$

implies  $\partial|\mathbf{A}|/\partial a_{ij} = A_{ij}$ . Thus, regarding  $|\mathbf{A} + a\mathbf{U}|$  as a function of  $a$ , setting  $a'_{ij} = a_{ij} + a$ , and denoting by  $A'_{ij}$  the cofactor of  $a'_{ij}$  in  $\mathbf{A} + a\mathbf{U}$ , we have

$$\frac{d}{da} |\mathbf{A} + a\mathbf{U}| = \sum_{i,j=1}^n \frac{\partial}{\partial a'_{ij}} |\mathbf{A} + a\mathbf{U}| \frac{d}{da} (a'_{ij}) = \sum_{i,j=1}^n A'_{ij}.$$

But since  $\sum_{j=1}^n A'_{ij} \cdot 1$  is evidently the expansion about the  $i$ th row of a matrix obtained from  $\mathbf{A} + a\mathbf{U}$  by replacing the  $i$ th row by  $1, 1, \dots, 1$ , that is, the matrix

$$\begin{vmatrix} a_{11} + a & a_{12} + a & \cdots & a_{1n} + a \\ \vdots & \vdots & & \vdots \\ a_{i-1,1} + a & a_{i-1,2} + a & \cdots & a_{i-1,n} + a \\ 1 & 1 & \cdots & 1 \\ a_{i+1,1} + a & a_{i+1,2} + a & \cdots & a_{i+1,n} + a \\ \vdots & \vdots & & \vdots \\ a_{n1} + a & a_{n2} + a & \cdots & a_{nn} + a \end{vmatrix},$$

we can subtract  $a$  times the  $i$ th row from all others to obtain

$$\sum_{j=1}^n A'_{ij} = \sum_{j=1}^n A_{ij}. \quad (\text{A.9.1})$$

Thus

$$\frac{d}{da} |\mathbf{A} + a\mathbf{U}| = \sum_{i,j=1}^n A_{ij},$$

which is independent of  $a$ . It follows that  $|\mathbf{A} + a\mathbf{U}|$  is the linear function

$$|\mathbf{A}| + a \cdot \sum_{i,j=1}^n A_{ij},$$

which is our first identity. Further, (A.9.1) states precisely that

$$\begin{vmatrix} A'_{11} & A'_{12} & \cdots & A'_{1n} \\ A'_{21} & A'_{22} & \cdots & A'_{2n} \\ \vdots & \vdots & & \vdots \\ A'_{n1} & A'_{n2} & \cdots & A'_{nn} \end{vmatrix} \begin{vmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{vmatrix} = \begin{vmatrix} \sum A'_{1j} \\ \sum A'_{2j} \\ \vdots \\ \sum A'_{nj} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{vmatrix} \begin{vmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{vmatrix},$$

that is, our third identity. The second follows by the original argument concerning

$$\sum_{j=1}^n A'_{ij} \quad \text{applied to} \quad \sum_{i=1}^n A'_{ij}$$

as a column expansion.

It follows directly from the definition that  $|\lambda \mathbf{A}| = \lambda^n |\mathbf{A}|$  for any  $n \times n$  matrix  $\mathbf{A}$ . Consequently, since

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \text{adj}(\mathbf{A})$$

when  $\mathbf{A}$  is nonsingular, we have

$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|} = \frac{1}{|\mathbf{A}|^n} |\text{adj}(\mathbf{A})|,$$

and thus  $|\text{adj}(\mathbf{A})| = |\mathbf{A}|^{n-1}$ . Since this last relation holds for any limit of nonsingular matrices, it holds for all  $\mathbf{A}$ .

For any nonsingular matrix  $\mathbf{A}$ ,

$$\begin{aligned} \mathbf{A}^{-1} \begin{pmatrix} i_1, i_2, \dots, i_p \\ j_1, j_2, \dots, j_p \end{pmatrix} \\ = \frac{\mathbf{A} \begin{pmatrix} j'_1, j'_2, \dots, j'_{n-p} \\ i'_1, i'_2, \dots, i'_{n-p} \end{pmatrix}}{|\mathbf{A}|} (-1)^{\sum_{\nu=1}^{p(i_p+j_p)}} \end{aligned} \quad (\text{A.9.2})$$

where  $\{j'_1, j'_2, \dots, j'_{n-p}\}$  forms the complementary set of column indices to  $\{j_1, \dots, j_p\}$ , and similarly for  $\{i'_1, \dots, i'_{n-p}\}$ . Equation (A.9.2) is known as Sylvester's identity.

Consider the case in which  $i_1 = j_1 = 1$ ,  $i_2 = j_2 = 2$ ,  $\dots$ ,  $i_p = j_p = p$ , so that the identity becomes

$$\mathbf{A}^{-1} \begin{pmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p \end{pmatrix} = \frac{1}{|\mathbf{A}|} \mathbf{A} \begin{pmatrix} p+1, p+2, \dots, n \\ p+1, p+2, \dots, n \end{pmatrix}.$$

Noting that the transpose of a matrix has the same determinant as the matrix and that  $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$ , and denoting by  $A_{ij}$  the cofactor of  $a_{ij}$ , we have the product

$$\begin{vmatrix}
 A_{11} & A_{12} & \cdots & A_{1p} & A_{1,p+1} & \cdots & \cdots & A_{1n} \\
 A_{21} & A_{22} & \cdots & A_{2p} & A_{2,p+1} & \cdots & \cdots & A_{2n} \\
 \vdots & & & & & & & \\
 A_{p1} & A_{p2} & \cdots & A_{pp} & A_{p,p+1} & \cdots & \cdots & A_{pn} \\
 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\
 \vdots & \vdots & & \vdots & \vdots & & & \vdots \\
 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 1
 \end{vmatrix} \times \begin{vmatrix}
 a_{11} & a_{21} & \cdots & a_{n1} \\
 a_{12} & a_{22} & \cdots & a_{n2} \\
 \vdots & & & \\
 a_{1p} & a_{2p} & \cdots & a_{n,p} \\
 a_{1,p+1} & a_{2,p+1} & \cdots & a_{n,p+1} \\
 \vdots & \vdots & & \vdots \\
 \vdots & \vdots & & \vdots \\
 a_{1n} & a_{2n} & \cdots & a_{nn}
 \end{vmatrix}$$

(where the first  $p$  rows of the first determinant coincide with those of  $\text{adj}(\mathbf{A})$ , while the last  $n - p$  feature only 1's on the main diagonal as non-zero entries), equal to

$$\begin{vmatrix}
 |\mathbf{A}| & 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\
 0 & |\mathbf{A}| & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 \vdots & & & & & & & & \\
 0 & \cdots & \cdots & \cdots & 0 & |\mathbf{A}| & 0 & \cdots & 0 \\
 a_{1,p+1} & \cdots & \cdots & \cdots & \cdots & a_{p,p+1} & a_{p+1,p+1} & \cdots & a_{n,p+1} \\
 a_{1,p+2} & \cdots & \cdots & \cdots & \cdots & a_{p,p+2} & a_{p+1,p+2} & \cdots & a_{n,p+2} \\
 \vdots & & & & & & & & \\
 a_{1,n} & \cdots & \cdots & \cdots & \cdots & a_{p,n} & a_{p+1,n} & \cdots & a_{n,n}
 \end{vmatrix} = |\mathbf{A}|^p \mathbf{A} \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix}$$

(expanding by rows, using the first  $p$  rows). On the other hand, if we expand the first factor of our product by rows (using the last  $n - p$  rows), it becomes, with  $\mathbf{C} = \text{adj}(\mathbf{A})$ ,

$$\mathbf{C} \begin{pmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p \end{pmatrix} |\mathbf{A}|,$$

hence

$$\mathbf{C} \begin{pmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p \end{pmatrix} = |\mathbf{A}|^{p-1} \mathbf{A} \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix}.$$

But

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \mathbf{C},$$

and thus

$$\mathbf{A}^{-1} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix} = \frac{1}{|\mathbf{A}|^p} \mathbf{C} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix} = \frac{1}{|\mathbf{A}|} \mathbf{A} \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix},$$

which is exactly the identity in our special case.

By further interchanging appropriate rows and columns we may achieve (A.9.2). In terms of  $\mathbf{C} = \text{adj}(\mathbf{A})$  the identity (A.9.2) becomes

$$\frac{1}{|\mathbf{A}|^p} \mathbf{C} \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix} = (-1)^{\sum_{\nu=1}^p (i_{\nu} + j_{\nu})} \frac{1}{|\mathbf{A}|} \mathbf{A} \begin{pmatrix} j'_1, \dots, j'_{n-p} \\ i'_1, \dots, i'_{n-p} \end{pmatrix},$$

or

$$\mathbf{C} \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix} = (-1)^{\sum_{\nu=1}^p (i_{\nu} + j_{\nu})} |\mathbf{A}|^{p-1} \mathbf{A} \begin{pmatrix} j'_1, \dots, j'_{n-p} \\ i'_1, \dots, i'_{n-p} \end{pmatrix},$$

since

$$\frac{1}{|\mathbf{A}|} \mathbf{C} = \mathbf{A}^{-1}.$$

The determinants

$$\mathbf{A} \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix}$$

are clearly generalized minors of  $\mathbf{A}$ ; but in particular the  $ij$ th minor of  $\mathbf{A}$  is

$$\mathbf{A} \begin{pmatrix} 1, 2, \dots, i-1, i+1, \dots, n \\ 1, 2, \dots, j-1, j+1, \dots, n \end{pmatrix},$$

while of course

$$\mathbf{A} \begin{pmatrix} 1, \dots, n \\ 1, \dots, n \end{pmatrix} = |\mathbf{A}|.$$

Let  $p < n$  be fixed, and set

$$d_{ij} = \mathbf{A} \begin{pmatrix} 1, \dots, p, i \\ 1, \dots, p, j \end{pmatrix} \quad (i, j \geq p+1),$$

so that the  $(n - p) \times (n - p)$  array  $\mathbf{D} = \|d_{ij}\|$  forms a matrix. Then (*Sylvester's theorem*)

$$\mathbf{D} \begin{pmatrix} i_1, \dots, i_q \\ j_1, \dots, j_q \end{pmatrix} = \mathbf{A} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix}^{q-1} \mathbf{A} \begin{pmatrix} 1, \dots, p, i_1, \dots, i_q \\ 1, \dots, p, j_1, \dots, j_q \end{pmatrix}. \quad (\text{A.9.3})$$

Since we shall be concerned only with the matrix obtained from  $\mathbf{A}$  by deleting all rows but  $1, 2, \dots, p, i_1, \dots, i_q$  and all columns but  $1, 2, \dots, p, j_1, \dots, j_q$ , we can assume without loss of generality that this matrix forms our original  $\mathbf{A}$  and thus prove the identity in the one case in which  $q = n - p$ :

$$\begin{aligned} \mathbf{D} \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix} &= \mathbf{A} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix}^{n-p-1} \mathbf{A} \begin{pmatrix} 1, \dots, n \\ 1, \dots, n \end{pmatrix} \\ &= \mathbf{A} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix}^{n-p-1} \cdot |\mathbf{A}|. \end{aligned}$$

Let  $\mathbf{C} = \text{adj}(\mathbf{A})$ . Then

$$c_{ij} = (-1)^{i+j} \mathbf{A} \begin{pmatrix} 1, \dots, j-1, j+1, \dots, n \\ 1, \dots, i-1, i+1, \dots, n \end{pmatrix}.$$

In view of our first identity (written in terms of the adjoint), we have

$$\begin{aligned} \mathbf{C} \begin{pmatrix} p+1, \dots, r-1, r+1, \dots, n \\ p+1, \dots, s-1, s+1, \dots, n \end{pmatrix} &= (-1)^{r+s} |\mathbf{A}|^{n-p-2} \mathbf{A} \begin{pmatrix} 1, \dots, p, s \\ 1, \dots, p, r \end{pmatrix} \quad (\text{A.9.4}) \\ &= (-1)^{r+s} |\mathbf{A}|^{n-p-2} d_{sr}, \end{aligned}$$

and

$$\mathbf{C} \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix} = |\mathbf{A}|^{n-p-1} \mathbf{A} \begin{pmatrix} 1, \dots, p \\ 1, \dots, p \end{pmatrix}. \quad (\text{A.9.5})$$

But

$$b_{sr} = (-1)^{s+r} \mathbf{C} \begin{pmatrix} p+1, \dots, r-1, r+1, \dots, n \\ p+1, \dots, s-1, s+1, \dots, n \end{pmatrix} \quad (r, s \geq p+1)$$

defines the elements of the adjoint  $\mathbf{B}$  of the  $(n - p) \times (n - p)$  matrix

$$\begin{vmatrix} c_{p+1,p+1} & c_{p+1,p+2} & \cdots & c_{p+1,n} \\ c_{p+2,p+2} & \cdots & \cdots & c_{p+2,n} \\ \vdots & & & \vdots \\ c_{n,p+1} & \cdots & \cdots & c_{n,n} \end{vmatrix},$$

and thus

$$\mathbf{B} \binom{p+1, \dots, n}{p+1, \dots, n} = |\mathbf{B}| = \mathbf{C} \binom{p+1, \dots, n}{p+1, \dots, n}^{n-p-1}$$

since

$$\mathbf{C} \binom{p+1, \dots, n}{p+1, \dots, n}$$

is just the determinant of this matrix and  $|\text{adj}(\mathbf{A}_0)| = |\mathbf{A}_0|^{k-1}$  for any  $k \times k$  matrix  $\mathbf{A}_0$ . But clearly  $b_{sr} = |\mathbf{A}|^{n-p-2} d_{sr}$ ; hence

$$|\mathbf{B}| = |\mathbf{A}|^{(n-p-2)(n-p)} \mathbf{D} \binom{p+1, \dots, n}{p+1, \dots, n}$$

and

$$\mathbf{C} \binom{p+1, \dots, n}{p+1, \dots, n}^{n-p-1} = |\mathbf{A}|^{(n-p)(n-p-2)} \mathbf{D} \binom{p+1, \dots, n}{p+1, \dots, n}.$$

Thus by (A.9.5)

$$|\mathbf{A}|^{(n-p-1)^2} \mathbf{A} \binom{1, \dots, p}{1, \dots, p}^{n-p-1} = |\mathbf{A}|^{(n-p)(n-p-2)} \mathbf{D} \binom{p+1, \dots, n}{p+1, \dots, n},$$

or

$$\mathbf{D} \binom{p+1, \dots, n}{p+1, \dots, n} = |\mathbf{A}| \mathbf{A} \binom{1, \dots, p}{1, \dots, p}^{n-p-1},$$

completing the proof.

**A.10 Compound matrices.** From an  $n \times n$  matrix  $\mathbf{A}$  we form the  $p$ th associated (compound) matrix  $\mathcal{C}_p$ , whose elements comprise all  $p \times p$  sub-determinants of  $\mathbf{A}$ , arranged in alphabetical order (viz.,  $\{1, 3\}$  precedes  $\{1, 8\}$ ;  $\{2, 6\}$  precedes  $\{3, 1\}$ ). For example, the following matrix  $\mathcal{C}_2$  is a square matrix of size  $\binom{n}{2} \times \binom{n}{2}$ :

$$\left\| \begin{array}{cccccc} A \begin{pmatrix} 1, 2 \\ 1, 2 \end{pmatrix} & A \begin{pmatrix} 1, 2 \\ 1, 3 \end{pmatrix} & \cdots & A \begin{pmatrix} 1, 2 \\ 1, n \end{pmatrix} & A \begin{pmatrix} 1, 2 \\ 2, 3 \end{pmatrix} & \cdots & A \begin{pmatrix} 1, 2 \\ n-1, n \end{pmatrix} \\ A \begin{pmatrix} 1, 3 \\ 1, 2 \end{pmatrix} & A \begin{pmatrix} 1, 3 \\ 1, 3 \end{pmatrix} & \cdots & A \begin{pmatrix} 1, 3 \\ 1, n \end{pmatrix} & A \begin{pmatrix} 1, 3 \\ 2, 3 \end{pmatrix} & \cdots & A \begin{pmatrix} 1, 3 \\ n-1, n \end{pmatrix} \\ \vdots & & & & & & \vdots \\ A \begin{pmatrix} n-1, n \\ 1, 2 \end{pmatrix} & \cdots & \cdots & \cdots & \cdots & \cdots & A \begin{pmatrix} n-1, n \\ n-1, n \end{pmatrix} \end{array} \right\|$$

Similarly,  $\mathfrak{Q}_p$  is a matrix of size  $\binom{n}{p} \times \binom{n}{p}$ .

It follows immediately from formula (A.8.3) that if  $\mathbf{A} \cdot \mathbf{B} = \mathbf{C}$ , then  $\mathfrak{Q}_p \cdot \mathfrak{G}_p = \mathfrak{C}_p$ , where  $\mathfrak{Q}_p$ ,  $\mathfrak{G}_p$ , and  $\mathfrak{C}_p$  are the  $p$ th compound matrices of  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , respectively. In particular, if  $\mathbf{B} = \mathbf{A}^{-1}$ , then  $\mathfrak{G}_p$  is the inverse matrix of  $\mathfrak{Q}_p$ .

Let  $\lambda_1, \lambda_2, \dots, \lambda_n$  denote a complete system of eigenvalues (allowing repetitions when required) of the matrix  $\mathbf{A}$ , and suppose that  $\mathbf{A} = \mathbf{PTP}^{-1}$ , where

$$\mathbf{T} = \left\| \begin{array}{cccccc} \lambda_1 & & & \circ & & & \\ \cdot & \lambda_2 & & & & & \\ \cdot & & \cdot & & & & \\ \cdot & & & \cdot & & & \\ \cdot & \cdot & & & \lambda_n & & \end{array} \right\|$$

represents the Jordan canonical form of  $\mathbf{A}$ . Then manifestly

$$\mathfrak{Q}_p = \mathfrak{G}_p \mathfrak{J}_p \mathfrak{G}_p^{-1}, \quad (\text{A.10.1})$$

where  $\mathfrak{J}_p$  is a triangular matrix whose diagonal elements are

$$\mathbf{T} \begin{pmatrix} i_1, \dots, i_p \\ i_1, \dots, i_p \end{pmatrix} = \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_p}.$$

Thus the eigenvectors of  $\mathfrak{Q}_p$  are apparent, being the diagonal entries of  $\mathfrak{J}_p$ .

In the important special case in which the eigenvalues are distinct, we have verified (p. 376) that the similarity transformation  $\mathbf{P}$  can be formed as a matrix whose column vectors are the eigenvectors of  $\lambda_1, \lambda_2, \dots, \lambda_n$ , respectively. But we see by inspection of (A.10.1) that since  $\mathfrak{J}_p$  is diagonal, the eigenvector corresponding to the eigenvalue  $\lambda_{k_1} \lambda_{k_2} \cdots \lambda_{k_p}$  is the vector

$$\mathbf{P} \begin{pmatrix} \alpha_1, \alpha_2, \dots, \alpha_p \\ k_1, k_r, \dots, k_p \end{pmatrix},$$

where  $(\alpha_1, \dots, \alpha_p)$  designates the variable component index.

A matrix  $\mathbf{A}$  is said to be totally positive (T.P.) of order  $p$  if

$$\mathbf{A} \begin{pmatrix} i_1, i_2, \dots, i_p \\ j_1, j_2, \dots, j_p \end{pmatrix} \geq 0$$

for every pair of index sets  $1 \leq i_1 < i_2 < \dots < i_p \leq n$  and  $1 \leq j_1 < j_2 < \dots < j_p \leq n$ . Because of (A.8.3) the product of two totally positive matrices of order  $p$  is again totally positive of order  $p$ . A matrix is said to be weakly totally positive if

$$(-1)^{\sum_{\nu=1}^p (i_\nu + j_\nu)} \mathbf{A} \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix} \geq 0 \quad \text{for} \\ 1 \leq \frac{i_1 < i_2 < \dots < i_p}{j_1 < j_2 < \dots < j_p} \leq n.$$

We may immediately deduce from Sylvester's identity (A.9.2) that if  $\mathbf{A}$  is T.P. of order  $p$ , then  $\mathbf{B} = \mathbf{A}^{-1}$  (provided the inverse exists) is weakly T.P. of order  $p$ .

By way of illustration, we suppose that  $\mathbf{A} = \|a_{ij}\|$ , where

$$a_{ij} = \begin{cases} b_i c_j & i \leq j, \\ b_j c_i & i \geq j, \end{cases}$$

and all  $b_i$  and  $c_i$  have the same sign and are nonzero. A calculation will show that  $\mathbf{A}$  is T.P. of all orders if and only if

$$\frac{b_1}{c_1} \leq \frac{b_2}{c_2} \leq \dots \leq \frac{b_n}{c_n}.$$

### EXERCISES

1. If  $\mathbf{A}$  is normal or unitary, show that the same is true for  $\mathbf{Q}_p$ .
2. Prove that if  $\mathbf{A}$  is symmetric,  $\mathbf{Q}_p$  is symmetric; and that if the quadratic form for  $\mathbf{A}$  is positive definite, the same holds for  $\mathbf{Q}_p$ .
3. Prove that if  $|\mathbf{A}| > 0$  and  $\mathbf{A}$  is weakly T.P. of order  $p$ , then  $\mathbf{A}^{-1}$  is T.P. of order  $p$ .
4. Find  $|\mathbf{Q}_p|$  in terms of  $|\mathbf{A}|$ .

## APPENDIX B

### CONVEX SETS AND CONVEX FUNCTIONS

**B.1 Convex sets in  $E^n$ .** A set  $X$  of vectors in  $E^n$  is called a convex set if, for any two vectors  $\mathbf{x}$  and  $\mathbf{y}$  in  $X$ , the following relations hold:  $\alpha\mathbf{x} + \beta\mathbf{y} \in X$ ;  $\alpha, \beta \geq 0$ ;  $\alpha + \beta = 1$ . Among the simpler convex sets are linear subspaces, spheres, and triangles.

For a finite number of vectors  $\mathbf{a}^1, \dots, \mathbf{a}^r$  and real numbers  $\alpha_1, \dots, \alpha_r$ , the set  $H$  defined by

$$H = \{\mathbf{x} | (\mathbf{a}^i, \mathbf{x}) = \alpha_i, \quad i = 1, \dots, r\}$$

is a closed convex set. In particular, for a nonzero vector  $\mathbf{a}$  and a real number  $\alpha$ , the set  $H = \{\mathbf{x} | (\mathbf{a}, \mathbf{x}) = \alpha\}$  is called a hyperplane [or an  $(n - 1)$ -dimensional linear variety].

A hyperplane  $H$  determines two *closed half-spaces*

$$H_1 = \{\mathbf{x} | (\mathbf{a}, \mathbf{x}) \geq \alpha\}, \quad H_2 = \{\mathbf{x} | (\mathbf{a}, \mathbf{x}) \leq \alpha\},$$

both of which are closed, convex sets.

A hyperplane  $H$  is said to be a *supporting hyperplane* to a convex set  $X$  if  $X$  is contained in one of the half-spaces of  $H$  and the boundary of  $X$  has a point in common with  $H$ . More precisely, a hyperplane  $\{\mathbf{a}, \alpha\}$  is a supporting hyperplane to  $X$  if

$$\inf_{\mathbf{x} \in X} (\mathbf{a}, \mathbf{x}) = \alpha.$$

(Here the appropriate infinite point must be admitted as a possible point.)

► **LEMMA B.1.1.** Let  $X$  be a convex set and  $\mathbf{y}$  be a point exterior to the closure of  $X$ . Then there exists a vector  $\mathbf{a}$  such that

$$\inf_{\mathbf{x} \in X} (\mathbf{a}, \mathbf{x}) > (\mathbf{a}, \mathbf{y}).$$

*Proof.* Let  $\mathbf{x}^0$  be a boundary point of  $X$  such that

$$\sqrt{(\mathbf{y} - \mathbf{x}^0, \mathbf{y} - \mathbf{x}^0)} = |\mathbf{y} - \mathbf{x}^0| = \inf_{\mathbf{x} \in X} |\mathbf{y} - \mathbf{x}|. \quad (\text{B.1.1})$$

If  $\mathbf{x} \in X$ , then  $\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0) \in \overline{X}$  ( $0 \leq t \leq 1$ ), where  $\overline{X}$  denotes the closure of  $X$ .

By (B.1.1)

$$|\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0) - \mathbf{y}|^2 \geq |\mathbf{x}^0 - \mathbf{y}|^2. \quad (\text{B.1.2})$$

Expanding (B.1.2), we obtain

$$2t(\mathbf{x} - \mathbf{x}^0, \mathbf{x}^0 - \mathbf{y}) + t^2(\mathbf{x} - \mathbf{x}^0, \mathbf{x} - \mathbf{x}^0) \geq 0 \quad (0 \leq t \leq 1).$$

Hence  $(\mathbf{x} - \mathbf{x}^0, \mathbf{x}^0 - \mathbf{y}) \geq 0$ , or

$$(\mathbf{x}, \mathbf{x}^0 - \mathbf{y}) \geq (\mathbf{x}^0, \mathbf{x}^0 - \mathbf{y}) > (\mathbf{y}, \mathbf{x}^0 - \mathbf{y}).$$

Let  $\mathbf{a} = \mathbf{x}^0 - \mathbf{y}$ . Then  $\mathbf{a} \neq \mathbf{0}$  and

$$\inf_{\mathbf{x} \in X} (\mathbf{x}, \mathbf{a}) \geq (\mathbf{x}^0, \mathbf{a}) > (\mathbf{y}, \mathbf{a}),$$

as asserted.

► **LEMMA B.1.2.** Let  $X$  be a convex set and  $\mathbf{y}$  be on the boundary of  $X$ . Then there exists a supporting plane through  $\mathbf{y}$ ; i.e., there exists a nonzero vector  $\mathbf{a}$  such that

$$\inf_{\mathbf{x} \in X} (\mathbf{a}, \mathbf{x}) = (\mathbf{a}, \mathbf{y}).$$

*Proof.* Consider a sequence  $\{\mathbf{y}^\nu\}$ , with  $\mathbf{y}^\nu$  exterior to the closure of  $X$  and such that  $\lim_\nu \mathbf{y}^\nu = \mathbf{y}$ .

By Lemma B.1.1, there is a sequence  $\{\mathbf{a}^\nu\}$  which may be normalized such that

$$(\mathbf{a}^\nu, \mathbf{a}^\nu) = 1, \quad \inf_{\mathbf{x} \in X} (\mathbf{x}, \mathbf{a}^\nu) > (\mathbf{y}^\nu, \mathbf{a}^\nu) \quad (\nu = 1, 2, \dots).$$

Taking a limiting point  $\mathbf{a}$  of  $\{\mathbf{a}^\nu\}$ , we have, for any  $\mathbf{x} \in X$ ,

$$(\mathbf{x}, \mathbf{a}) = \lim_\nu (\mathbf{x}, \mathbf{a}^\nu) \geq \lim_\nu (\mathbf{y}^\nu, \mathbf{a}^\nu) = (\mathbf{y}, \mathbf{a}),$$

as asserted.

The preceding two lemmas in conjunction yield the following theorem.

► **THEOREM B.1.1.** A closed convex set is the intersection of all its supporting half-spaces, and every boundary point of the set lies on a supporting hyperplane.

► **THEOREM B.1.2.** If  $X$  and  $Y$  are two convex sets with no interior point in common, then there is a hyperplane  $H$  that separates  $X$  and  $Y$ ; that is, there exist a nonzero vector  $\mathbf{a}$  and a scalar  $\alpha$  such that  $(\mathbf{a}, \mathbf{x}) \geq \alpha$  for all  $\mathbf{x} \in X$  and  $(\mathbf{a}, \mathbf{y}) \leq \alpha$  for all  $\mathbf{y} \in Y$ .

*Proof.* Consider the set

$$X - Y = \{\mathbf{x} - \mathbf{y} | \mathbf{x} \in X, \mathbf{y} \in Y\},$$

which is convex and does not contain  $\mathbf{0}$  in the interior. By Lemmas B.1.1 and B.1.2 such a set has a vector  $\mathbf{a}$  ( $\mathbf{a} \neq \mathbf{0}$ ) such that

$$(\mathbf{a}, \mathbf{x} - \mathbf{y}) \geq (\mathbf{a}, \mathbf{0}) = 0 \quad (\mathbf{x} \in X, \mathbf{y} \in Y).$$

The assertion of the theorem follows immediately.

► **THEOREM B.1.3.** If  $X$  and  $Y$  are closed convex sets having no point in common and at least one of them is bounded, then there exists a hyperplane  $H$  determined by the parameters  $\{\mathbf{a}, \alpha\}$  that strictly separates  $X$  and  $Y$ : that is,  $(\mathbf{a}, \mathbf{x}) > \alpha$  for all  $\mathbf{x} \in X$  and  $(\mathbf{a}, \mathbf{y}) < \alpha$  for all  $\mathbf{y} \in Y$ .

*Proof.* In this case  $X - Y$  is closed and does not contain  $\mathbf{0}$ ; hence the theorem follows immediately from Lemma B.1.1.

The following result is frequently used in the text.

► **LEMMA B.1.3.** Let  $X$  be a convex set which has no point in common with the nonnegative orthant. Then there exists a vector  $\mathbf{a} > \mathbf{0}$  (every component of  $\mathbf{a}$  is nonnegative and at least one component is positive) such that  $(\mathbf{a}, \mathbf{x}) \leq 0$  for all  $\mathbf{x} \in X$ .

*Proof.* Apply Theorem B.1.2 to  $X$  and  $Y$ , where  $Y$  denotes the nonnegative orthant. Then there exists a vector  $\mathbf{a} \neq \mathbf{0}$  such that  $(\mathbf{a}, \mathbf{x}) \leq 0$  for all  $\mathbf{x} \in X$  and  $(\mathbf{a}, \mathbf{y}) \geq 0$  for all  $\mathbf{y} \in Y$ . It is easy to show that  $\mathbf{a} > \mathbf{0}$ .

### EXERCISES

- Theorem B.1.3 is not necessarily true if both  $X$  and  $Y$  are unbounded. Show that a counterexample can be constructed for  $E^2 = \{\langle x_1, x_2 \rangle\}$  by taking  $X = \{\langle x_1, x_2 \rangle | x_1 \leq 0\}$   $Y = \{\langle x_1, x_2 \rangle | x_2 \geq e^{-x_1}\}$ .
- Prove that every set and its closure have the same exterior points. Show also that for convex sets it is also true that every convex set and its closure have the same inner points, hence also the same boundary points.
- Prove that in  $E^2$  every polygon is convex if and only if its interior angles are less than 180 degrees.
- Prove that the intersection of any number of convex sets is convex. Prove that the closure of a convex set is convex.
- Prove that every set in  $E^n$  can be embedded uniquely in a subspace of lowest possible dimension.
- To obtain the convex hull of a closed bounded set in  $E^2$  we can “wrap a string around the set and pull it tight.” The area determined by the string and its interior is the convex hull. Show that this method works for finding the convex hull of two triangles, or two circles. What does Lemma B.2.2 say about the convex hull of two triangles (or circles)?

**B.2 Convex hulls of sets and extreme points of convex sets.** The smallest convex set containing a set  $X$  is called the *convex hull* of  $X$ . It is usually

denoted by  $[X]$ , sometimes by  $\text{Co}(X)$ . The set  $[X]$  may be constructively formed as follows:

$$[X] = \left\{ \sum_{k=1}^r \alpha_k \mathbf{x}^k \mid \mathbf{x}^k \in X, \sum_{k=1}^r \alpha_k = 1, \alpha_k \geq 0, k = 1, \dots, r \right\},$$

where  $r$  is an arbitrary positive integer. If in particular  $X = \{\mathbf{a}^1, \dots, \mathbf{a}^r\}$ , then  $[\mathbf{a}^1, \dots, \mathbf{a}^r]$  is referred to as the polyhedral convex set spanned by  $\mathbf{a}^i$ . It is easily shown that if  $X$  is bounded,  $[X]$  is bounded, and that if  $X$  is compact,  $[X]$  is compact. The set  $[\mathbf{a}^1, \dots, \mathbf{a}^r]$  is clearly closed and bounded.

The following well-known representation is of fundamental value.

►LEMMA B.2.1. Let  $X$  be a subset of  $E^n$ , and let  $[X]$  be the convex hull of  $X$ . Then every point of  $[X]$  can be represented as a convex combination of at most  $n + 1$  points of  $X$ .

*Proof.* It suffices to show that if

$$\mathbf{x} = \sum_{k=1}^r \alpha_k \mathbf{x}^k \quad (\mathbf{x}^k \in X, \alpha_k > 0), \quad \text{and} \quad \sum_{k=1}^r \alpha_k = 1 \quad (r > n + 1),$$

then the number of vectors  $\mathbf{x}^1, \dots, \mathbf{x}^r$  of  $X$  used in the representation of  $\mathbf{x}$  can be reduced. Since  $\mathbf{x}^1, \dots, \mathbf{x}^r$  ( $r > n + 1$ ) are linearly dependent, there exist real numbers  $\beta_1, \dots, \beta_r$  not all zero such that

$$\sum_{k=1}^r \beta_k \mathbf{x}^k = \mathbf{0}, \quad \sum_{k=1}^r \beta_k = 0.$$

Let  $\epsilon$  be a real number such that  $\alpha_k + \epsilon\beta_k \geq 0$  ( $k = 1, \dots, r$ ) and  $\alpha_k + \epsilon\beta_k = 0$  for some  $k$ , say  $k_0$ . Then

$$\mathbf{x} = \mathbf{x} + \sum_{k=1}^r \epsilon\beta_k \mathbf{x}^k = \sum_{k=1}^r (\alpha_k + \epsilon\beta_k) \mathbf{x}^k,$$

where  $\alpha_k + \epsilon\beta_k \geq 0$  ( $k = 1, \dots, r$ ),  $\alpha_{k_0} + \epsilon\beta_{k_0} = 0$ ; and

$$\sum_{k=1}^r (\alpha_k + \epsilon\beta_k) = 1,$$

as was to be proved.

By imposing further restrictions on  $X$  we may sharpen the preceding result as follows.

►LEMMA B.2.2. If  $X$  has at most  $n$  connected components, then the number  $n + 1$  in Lemma B.2.1 may be decreased to  $n$ .

*Proof.* Consider the set  $S$  spanned by a fixed set of  $n + 1$  points of  $X$ , and assume that  $S$  has an interior point  $\mathbf{x}^0$  that cannot be represented by

a combination of  $n$  or fewer points of  $X$ . Corresponding to each  $(n - 2)$ -dimensional face  $L$  of  $S$ , let

$$T_L = \{\mathbf{x} = (1 - \lambda)\mathbf{x}^0 + \lambda\mathbf{y} \mid \lambda \leq 0, \mathbf{y} \in L\}.$$

The set  $T_L$  cannot intersect  $X$ , because if  $\mathbf{c} = (1 - \lambda)\mathbf{x}^0 + \lambda\mathbf{y}$  were in  $X$ , then  $\lambda \leq 0$  and

$$\mathbf{x}^0 = \frac{1}{1 - \lambda} \mathbf{c} + \frac{-\lambda}{1 - \lambda} \mathbf{y}$$

would represent  $\mathbf{x}^0$  as a combination of  $n$  or fewer points of  $X$ . The sets  $T_L$  obtained by traversing the  $(n - 2)$ -dimensional faces of  $S$  form the boundaries of  $n + 1$  nonoverlapping cones whose vertices are at  $\mathbf{x}^0$  and whose union is all of  $E^n$ . Furthermore, each vertex  $\mathbf{x}^i$  of  $S$  belongs to a different cone. Since  $X$  cannot intersect the boundaries of these cones,  $X$  must have at least  $n + 1$  connected components, thus contradicting the hypothesis.

► **LEMMA B.2.3.** If  $X$  is a bounded set in  $E^n$  and

$$\mathbf{x} = \sum_{i=1}^{\infty} \alpha_i \mathbf{x}^i,$$

where

$$\mathbf{x}^i \in X, \alpha_i > 0, \sum_{i=1}^{\infty} \alpha_i = 1,$$

then  $\mathbf{x}$  belongs to the convex hull of  $X$ .

This is easily proved by employing Theorem B.1.1 plus an induction argument based on the dimension of the smallest linear space containing the point  $\mathbf{x}$ . Suppose that the result has been proved for sets in  $E^n$ . Then let  $X$  be a bounded set in  $E^{n+1}$ . If  $\mathbf{x}$  is a boundary point of the closure of the convex span  $\Gamma$  of  $\{\mathbf{x}^i\}$ , then any supporting plane  $p$  to  $\Gamma$  at  $\mathbf{x}$  necessarily contains all  $\mathbf{x}^i$ . But the part common to  $p$  and  $\Gamma$  lies in a Euclidian space of dimension at most  $n$ , and therefore the induction hypothesis applies. Finally, if  $\mathbf{x}$  is interior to  $\Gamma$ , the result is immediate.

A point  $\mathbf{x}$  in  $X$  is called an *extreme point* of  $X$  if there are no points  $\mathbf{x}^1$  and  $\mathbf{x}^2$  in  $X$  such that  $\mathbf{x} = \lambda\mathbf{x}^1 + (1 - \lambda)\mathbf{x}^2$  for some  $\lambda$  ( $0 < \lambda < 1$ ), where  $\mathbf{x}^1 \neq \mathbf{x}^2$ .

► **LEMMA B.2.4.** A closed, bounded, convex set  $X$  is spanned by its extreme points. That is, every  $\mathbf{x}$  in  $X$  can be represented in the form

$$\mathbf{x} = \sum_{k=1}^r \lambda_k \mathbf{x}^k \quad (\lambda_k \geq 0; k = 1, \dots, r; \sum_k \lambda_k = 1),$$

where  $\mathbf{x}^1, \dots, \mathbf{x}^r$  are extreme points of  $X$ .

The proof again follows by an induction argument on the dimension of the smallest linear space containing  $X$ . We omit the formal details.

### EXERCISES

1. Let  $S$  be a convex subset of  $E^n$  containing the origin and  $n$  linearly independent vectors. Prove that  $S$  has inner points relative to  $E^n$ .
2. Apply Lemma B.2.1 to  $E^2$  to show that the convex hull of a set  $S$  is obtained by taking the union of all possible triangles in  $E^2$  with vertices in  $S$ .
3. Let  $e$  be the point in the closed bounded convex set  $S$  which is farthest from the origin. Prove that  $e$  is an extreme point of  $S$ .
4. Show that any closed convex set in  $E^n$  is the intersection of a denumerable number of closed half-spaces.
5. Prove that a closed convex set in  $E^n$  has a finite number of extreme points if and only if it is the intersection of a finite number of closed half-spaces.
6. Establish that a closed, bounded, convex set has extreme points in every supporting hyperplane.

**B.3 Convex cones.** A set  $X$  in  $E^n$  is called a *convex cone* if, for any  $x$  and  $y$  in  $X$ ,

$$\alpha x + \beta y \in X \quad (\alpha, \beta \geq 0).$$

A convex cone, of course, is a *convex set*. Linear subspaces and half-spaces are examples of convex cones.

The smallest convex cone that contains a set  $X$  is called the *convex cone spanned by  $X$*  and is denoted by  $\mathcal{P}(X)$ . Equivalently,

$$\mathcal{P}(X) = \left\{ \sum_{k=1}^r \alpha_k x^k \mid \alpha_k \geq 0, x^k \in X, k = 1, \dots, r; r \text{ arbitrary} \right\}.$$

The polar cone  $X^+$  is defined as the set of all directed separating hyperplanes to a set  $X$  which contain the origin. In symbols,

$$X^+ = \{a \mid (a, x) \geq 0 \quad \text{for all } x \in X\}. \quad (\text{B.3.1})$$

Clearly  $X^+$  is a closed convex cone.

For any sets  $X$  and  $Y$ , the following four properties are simple consequences of the definition (B.3.1) of polar cones.

- (i)  $X \subset Y$  implies  $X^+ \supset Y^+$ ;
- (ii)  $X \subset X^{++}$ ;
- (iii)  $X^+ = X^{+++}$ ;
- (iv)  $(X + Y)^+ = X^+ \cap Y^+ \quad \text{if } \mathbf{0} \in X, Y.$

► **THEOREM B.3.1.** For any closed convex cones  $X$  and  $Y$

- (I)  $X = X^{++}$  (duality),
- (II)  $(X + Y)^+ = X^+ \cap Y^+, \quad \left. \begin{array}{l} \\ \end{array} \right\}$  (modularity).
- (III)  $(X \cap Y)^+ = \overline{X^+ + Y^+}, \quad \left. \begin{array}{l} \\ \end{array} \right\}$  (modularity).

(Recall that a bar symbol over a set denotes closure.)

*Proof.* (I) Let  $\mathbf{y} \notin X$ . Then by Theorem B.1.3 there exists a vector  $\mathbf{a}$  such that  $(\mathbf{a}, \mathbf{x}) > (\mathbf{a}, \mathbf{y})$  for all  $\mathbf{x} \in X$ . Since  $X$  is a cone,  $(\mathbf{a}, \mathbf{x}) \geq 0 > (\mathbf{a}, \mathbf{y})$  for all  $\mathbf{x} \in X$ . Hence  $\mathbf{a} \in X^+$ ,  $\mathbf{y} \notin X^{++}$ , as was to be shown.

(II) This assertion was noted in property (iv) above.

$$(III) (X \cap Y)^+ = (X^{++} \cap Y^{++})^+ = (X^+ + Y^+)^{++} = \overline{X^+ + Y^+}.$$

The following theorem can be proved in a similar manner.

► **THEOREM B.3.2.** For any set  $X$ ,

$$\overline{\mathcal{P}(X)} = X^{++}.$$

*Convex polyhedral cones.* The convex cone spanned by a finite set of vectors  $\mathbf{a}^1, \dots, \mathbf{a}^r$  is called a *convex polyhedral cone* and denoted by

$$\mathcal{P}(\mathbf{a}^1, \dots, \mathbf{a}^r) = \left\{ \sum_{k=1}^r \lambda_k \mathbf{a}^k \mid \lambda_k \geq 0, k = 1, \dots, r \right\}.$$

Using matrix notation, we may write

$$\mathcal{P}(\mathbf{a}^1, \dots, \mathbf{a}^r) = \mathcal{P}(\mathbf{A}),$$

where  $\mathbf{A}$  is the  $n \times r$  matrix whose column vectors are  $\mathbf{a}^1, \dots, \mathbf{a}^r$ . In particular, for a nonzero vector  $\mathbf{a}$ , we have  $\mathcal{P}(\mathbf{a}) = \{\lambda \mathbf{a} \mid \lambda \geq 0\}$ . The set  $\mathcal{P}(\mathbf{a})$  is referred to as the *ray* spanned by  $\mathbf{a}$ .

A convex polyhedral cone is closed. This is a consequence of the following more general theorem.

► **THEOREM B.3.3.** A convex polyhedral cone is represented as the intersection of a finite set of supporting half-spaces. Equivalently, for any finite set of vectors  $\mathbf{a}^1, \dots, \mathbf{a}^r$ , there exist vectors  $\mathbf{b}^1, \dots, \mathbf{b}^s$  such that

$$\mathcal{P}(\mathbf{a}^1, \dots, \mathbf{a}^r) = (\mathbf{b}^1)^+ \cap \dots \cap (\mathbf{b}^s)^+,$$

and conversely.

The vectors  $\mathbf{b}^j$  may be found by taking all supporting planes passing through the origin to the convex space spanned by the vectors  $\mathbf{a}^i$ .

The duality theorem for convex polyhedral cones may be expressed in matrix notation as follows.

► **THEOREM B.3.4.** The inner product  $(\mathbf{x}, \mathbf{y})$  is nonnegative for all  $\mathbf{y} \geq \mathbf{0}$  if and only if  $\mathbf{x} = \mathbf{A}\mathbf{u}$  for some  $\mathbf{u} \geq \mathbf{0}$ .

The following theorem is useful in our study of linear programming.

► **THEOREM B.3.5.** Let  $X$  be a bounded convex polyhedral set or a convex closed cone having no point but  $\mathbf{0}$  in common with the nonnegative orthant of  $E^n$ . Then there exists a vector  $\mathbf{a}$  with strictly positive components  $(\mathbf{a} \gg \mathbf{0})$  such that  $(\mathbf{a}, \mathbf{x}) \leq 0$  for all  $\mathbf{x} \in X$ .

*Proof.* It is enough to prove the case when  $X$  is a cone. Let us suppose that  $\mathbf{a} \geq \mathbf{0}$  and  $(\mathbf{a}, \mathbf{x}) \leq 0$  for all  $\mathbf{x} \in X$  imply that  $a_i = 0$  for some component index  $i$ , say  $i = 1$ . Then

$$(-X)^+ \cap (I)^+ \subset H = \{\langle u_1, \dots, u_n \rangle | u_1 = 0\},$$

where  $I$  represents the nonnegative orthant. By the duality theorem (Theorem B.3.1),

$$\overline{-X + I} \supset H^+,$$

and it is readily shown that  $-X + I$  is already closed.

Since  $\langle \alpha, 0, 0, \dots, 0 \rangle \in H^+$  for all  $\alpha$ , there exist vectors  $\mathbf{x} \in X$  and  $\mathbf{b} \geq \mathbf{0}$  such that

$$-\mathbf{x} + \mathbf{b} = \langle \alpha, 0, \dots, 0 \rangle = \mathbf{a}.$$

If we take  $\alpha < 0$ , it is clear that  $\mathbf{x} = \mathbf{b} - \mathbf{a}$  is a nonnegative vector belonging to  $X$  and with a positive first component, in contradiction to the hypothesis of the theorem.

#### EXERCISES

1. In  $E^3$  let  $X = \{(x, y, 1) | x^2 + y^2 = 1\}$ . Find  $\mathcal{P}(x)$ .
2. In  $E^2$  let  $X = \{(x_1, x_2) | x_1 = \sqrt{x_2^2 + 1}\}$ . Then  $\mathcal{P}(X) = \{(0, 0)\} + \{(x_1, x_2) | 0 < |x_2| < x_1\}$ . Thus  $X$  and its convex hull may be closed without the same being true of  $\mathcal{P}(X)$ . Verify this statement.
3. Prove that  $\mathcal{P}(X) \supset \mathcal{P}(\overline{X})$ .
4. Let  $C$  be a convex set. Show that  $\mathcal{P}(C)$  consists of the union of all rays spanned by points of  $C$ .
5. Prove that if  $C$  is a closed bounded convex set,  $\mathcal{P}(C)$  is the convex hull of the union of all rays spanned by extreme points of  $C$ .
6. Let  $\mathbf{x}_0$  be a vector in  $X$ , a bounded convex polyhedral set. Prove that there exists a vector  $\mathbf{a} \gg \mathbf{0}$  such that  $(\mathbf{a}, \mathbf{x}) \leq (\mathbf{a}, \mathbf{x}_0)$  for all  $\mathbf{x}$  in  $X$ , if and only if there is no vector  $\mathbf{x}_1$  in  $X$  such that  $\mathbf{x}_1 - \mathbf{x}_0 > \mathbf{0}$ .

**B.4 Convex and concave functions.** A function  $f(\mathbf{x})$  defined on a convex set  $X$  of  $E^n$  is called a *convex* function if

$$f[t\mathbf{x} + (1 - t)\mathbf{y}] \leq tf(\mathbf{x}) + (1 - t)f(\mathbf{y}) \quad (\mathbf{x}, \mathbf{y} \in X; 0 \leq t \leq 1). \quad (\text{B.4.1})$$

If (B.4.1) holds with strict inequality whenever  $\mathbf{x} \neq \mathbf{y}$  and  $0 < t < 1$ , then  $f(\mathbf{x})$  is called *strictly convex*. A function  $f(\mathbf{x})$  is called *concave* (*strictly concave*) if  $-f(\mathbf{x})$  is convex (strictly convex).

Examples of convex functions are such functions as  $e^x$  and  $x_1^2 + x_2^2$ . The condition (B.4.1) implies that

$$f\left(\sum_{k=1}^r t_k \mathbf{x}^k\right) \leq \sum_{k=1}^r t_k f(\mathbf{x}^k), \quad (\text{B.4.2})$$

where  $\mathbf{x}^k \in X$ ,  $\sum t_k = 1$ ,  $t_k \geq 0$ ,  $k = 1, \dots, r$ .

We list the following important facts concerning convex functions, each of which is proved in a straightforward fashion.

(i) Let

$$[X, f] = \left\{ \begin{pmatrix} \mathbf{x} \\ \xi \end{pmatrix} \mid \mathbf{x} \in X, \xi \geq f(\mathbf{x}) \right\}.$$

Then  $f(\mathbf{x})$  is a convex function defined on  $X$  if and only if  $[X, f]$  is a convex set in  $E^{n+1}$ .

If  $f$  is strictly convex, then  $[X^0, f]$  is a strictly convex set  $S$ ; that is, the midpoint of any two points of the set  $S$  lies in the relative interior of  $S$ . Here  $X^0$  denotes the interior of  $X$ .

(ii) If  $f_k(\mathbf{x})$  ( $k = 1, \dots, r$ ) are convex functions defined on  $X$ , and  $\lambda_k \geq 0$  ( $k = 1, \dots, r$ ), then  $\sum_k \lambda_k f_k(\mathbf{x})$  is also a convex function on  $X$ .

(iii) Let  $f_v(\mathbf{x})$  be a set of convex functions on  $X$ ; then

$$X_0 = \{ \mathbf{x} \mid \mathbf{x} \in X, \sup_v f_v(\mathbf{x}) < \infty \}$$

is a convex set, and  $\sup_v f_v(\mathbf{x})$  is convex on  $X_0$ .

(iv) If  $f(\mathbf{x})$  is strictly convex on  $X$ , then  $f(\mathbf{x})$  has at most one local minimum. Any local minimum is an absolute minimum.

(v) If  $f(\mathbf{x})$  is convex on  $X$ , it is continuous in the relative interior of  $X$ .

(vi) For a boundary point  $\mathbf{y}$  of  $X$ ,

$$\varliminf_{\mathbf{x} \rightarrow \mathbf{y}} f(\mathbf{x}) \leq f(\mathbf{y}).$$

(vii) If  $X$  is an open convex set and  $f(\mathbf{x})$  is differentiable in  $X$ , we have another characterization of the convexity of  $f(\mathbf{x})$ . Specifically,  $f(\mathbf{x})$  is convex if and only if, for  $\mathbf{x}^0$  and  $\mathbf{x}$  in  $X$ ,

$$f(\mathbf{x}) - f(\mathbf{x}^0) \geq \sum_i \left( \frac{\partial f}{\partial x_i} \right)^0 (x_i - x_i^0).$$

(viii) If  $f(\mathbf{x})$  is convex on an open convex set  $X$ , then  $f(\mathbf{x})$  has second-order partial derivatives existing *almost everywhere* (in the sense of the Lebesgue measure).

(ix) If  $X$  is one-dimensional and  $f$  is convex on  $X$ , then  $f$  has a derivative except for at most a countable set of points. Moreover, the right-hand and left-hand derivatives exist everywhere and are right and left continuous, respectively.

(x) If  $f(\mathbf{x})$  is a twice-differentiable function, and  $X$  is a general convex set in  $E^n$ , then  $f(\mathbf{x})$  is convex if and only if the matrix of second derivatives

$$\left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j}$$

is positive semidefinite for every  $x$  in  $X$ . If

$$\left( \frac{\partial f}{\partial x_i \partial x_j} \right)_{i,j}$$

is positive definite, then  $f(\mathbf{x})$  is strictly convex.

### EXERCISES

1. Consider the following examples: Prove that  $(\mathbf{a}, \mathbf{x})$  is a convex function on  $E^n$ , but never strictly convex. For  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$  in  $E^n$ ,  $f(\mathbf{x}) = \sum_1^r \lambda_i |x_i|^{\alpha_i}$  ( $\alpha_i \geq 1$ ) is convex on  $E^n$  if all  $\lambda_i \geq 0$ .
2. Let  $f(\mathbf{x})$  be a convex function on  $E^n$ . Show by applying the theorem on supporting hyperplanes that we obtain the result that for any point  $\mathbf{x}_0$  in  $E^n$  there is a linear function  $l(\mathbf{x})$  (that is,  $(\mathbf{l}, \mathbf{x})$ ) such that  $l(\mathbf{x}) \leq f(\mathbf{x})$  and  $l(\mathbf{x}_0) = f(\mathbf{x}_0)$ . Compare this with (vii).

3. If  $f_k(x)$ ,  $k = 1, 2, \dots$  are nonnegative monotonically increasing convex functions on  $E^1$ , so is  $\Pi_{k=1}^n f_k(x)$ .

4. If in  $E^n$ ,  $f(\mathbf{x})$  is convex (strictly convex) and  $\mathbf{A} \not\equiv \mathbf{0}$ , then so is  $f(\mathbf{Ax} + \mathbf{b})$ , where  $\mathbf{A}$  and  $\mathbf{b}$  are a matrix and a vector of appropriate dimensions.

5. Let  $f(\mathbf{x})$  be a continuous function on a convex set  $X$  of  $E^n$ . Then  $f$  is convex if and only if

$$f\left(\frac{\mathbf{x} + \mathbf{y}}{2}\right) \leq \frac{f(\mathbf{x}) + f(\mathbf{y})}{2} \quad \mathbf{x}, \mathbf{y} \in X.$$

6. Construct a convex function on  $[0, 1]$  whose derivative fails to exist for an infinity of points.

## APPENDIX C

### MISCELLANEOUS TOPICS

**C.1 Semicontinuous and equicontinuous functions.** *Semicontinuous functions.* A real-valued function  $f(\mathbf{x})$  defined on a set  $X$  in  $E^n$  is said to be continuous at  $\mathbf{x}_0$  in  $X$  if whenever  $\mathbf{x}_n \rightarrow \mathbf{x}_0$  (i.e., whenever the Euclidian distance  $d(\mathbf{x}_n, \mathbf{x}_0)$  tends to zero),  $f(\mathbf{x}_n) \rightarrow f(\mathbf{x}_0)$ . An equivalent definition expressed in terms of neighborhoods runs as follows: For any  $\epsilon > 0$  there exists some  $\delta(\epsilon, \mathbf{x}_0) > 0$  such that if  $d(\mathbf{x}, \mathbf{x}_0) < \delta$ , then

$$-\epsilon \leq f(\mathbf{x}) - f(\mathbf{x}_0) \leq \epsilon. \quad (\text{C.1.1})$$

A function is continuous in  $X$  if it is continuous at every point of  $X$ .

If  $\delta$  may be chosen independent of  $\mathbf{x}_0$  in  $X$ , then  $f$  is said to be uniformly continuous in  $X$ .

A function  $f$  is said to be upper semicontinuous at  $\mathbf{x}_0$  if the right-hand inequality of (C.1.1) holds for any arbitrary prescribed  $\epsilon$  ( $\epsilon > 0$ ) and all values  $\mathbf{x}$  satisfying  $d(\mathbf{x}, \mathbf{x}_0) \leq \delta$  for a suitable positive  $\delta$ . Similarly,  $f$  is lower semicontinuous at  $\mathbf{x}_0$  if the left-hand inequality of (C.1.1) holds for any prescribed  $\epsilon > 0$  and the corresponding  $\delta$ . Equivalently,  $f$  is upper (lower) semicontinuous at  $\mathbf{x}_0$  if and only if

$$\overline{\lim}_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \leq f(\mathbf{x}_0) \left( \underline{\lim}_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \geq f(\mathbf{x}_0) \right).$$

If  $f$  is upper semicontinuous on  $X$  (i.e., at every point of  $X$ ), then the set  $L = \{\mathbf{x} | f(\mathbf{x}) < \alpha\}$  for each real  $\alpha$  is relatively open in  $X$ , and conversely. For if  $\mathbf{x}_0$  belongs to  $L$ , then there exists an  $\epsilon$  such that  $f(\mathbf{x}_0) + \epsilon < \alpha$ . By the definition of an upper semicontinuous function, there exists about  $\mathbf{x}_0$  an open sphere of radius  $\delta$  all of whose members, also in  $X$ , belong to  $L$ ; hence  $L$  is relatively open. The converse is proved similarly. Analogously, a function  $f$  is lower semicontinuous on  $X$  if  $\{\mathbf{x} | f(\mathbf{x}) > \alpha\}$  is relatively open for each real  $\alpha$ .

With the aid of these characterizations we may readily establish the following properties:

- (i) The set of all upper (lower) semicontinuous functions is closed under addition. The negative of an upper semicontinuous function is lower semicontinuous, and conversely.
- (ii) If  $f_\alpha$  are upper (lower) semicontinuous on  $X$  and bounded below (bounded above), then  $\inf_\alpha f_\alpha$  ( $\sup_\alpha f_\alpha$ ) is upper (lower) semicontinuous. (Here the last equivalent definition of upper semicontinuity applies most easily.)

(iii) An upper (lower) semicontinuous function defined on a compact set achieves its maximum (minimum).

(iv) An upper (lower) semicontinuous function defined on a compact set may be approached by a decreasing (increasing) sequence of continuous functions, and conversely.

*Equicontinuous functions.* A family of functions  $f_\alpha$  defined on  $X$  is said to be equicontinuous at  $\mathbf{x}_0$  if for any  $\epsilon > 0$  there exists some  $\delta(\epsilon, \mathbf{x}_0) > 0$  (independent of  $\alpha$ ) such that  $|f_\alpha(\mathbf{x}) - f_\alpha(\mathbf{x}_0)| < \epsilon$  for  $d(\mathbf{x}, \mathbf{x}_0) \leq \delta$  and all  $\alpha$ . The functions  $f_\alpha$  are said to be equicontinuous on  $X$  if they are equicontinuous at each point of  $X$ . When  $\delta$  may be chosen independent of  $\mathbf{x}_0$  in  $X$ , we speak of uniform equicontinuity for  $f_\alpha$ . To illustrate this concept, we exhibit two families of equicontinuous functions.

(1) Let  $K(s, t)$  be a uniformly continuous function defined on the unit square, and let  $f_\alpha(t)$  constitute a family of uniformly bounded functions ( $0 \leq t \leq 1$ ). Then

$$g_\alpha(s) = \int_0^1 K(s, t)f_\alpha(t) dt$$

represents a uniform equicontinuous family of functions. The estimate

$$|g_\alpha(s_1) - g_\alpha(s_2)| \leq M \max_{0 \leq t \leq 1} |K(s_1, t) - K(s_2, t)|,$$

where  $M$  is the bound of  $f_\alpha$ , clearly implies the assertion.

(2) Let  $f_\alpha(z)$  constitute a family of uniformly bounded functions that are analytic in the unit circle  $C$  of the complex plane. Then on  $X$  (any interior circle of  $C$ ) the functions  $f_\alpha$  are uniformly equicontinuous. We prove this statement by the same reasoning as in Example (1), using the Cauchy integral formula to represent  $f_\alpha$  for points of  $X$ .

For our purposes the most important property of equicontinuous functions is given by the classical Ascoli theorem, which may be stated as follows: *Let  $f_\alpha$  be a uniformly bounded equicontinuous family of real functions defined on a compact separable set  $X$ ; then we can select a subsequence  $f_{\alpha_i}$  which converges uniformly.*

We sketch the idea of the proof. We specify a countable dense subset  $\{\mathbf{x}_i\}$  of  $X$ . By the diagonal procedure we may select a subsequence  $f_{\alpha_i}$  which converges for all  $\mathbf{x}_i$ . Owing to the equicontinuity, it follows that  $f_{\alpha_i}$  converges uniformly as desired.

**C.2 Fixed-point theorems.** The following theorems are reviewed with no proofs included. For detailed discussion of these results, see [130].

*Brouwer's fixed-point theorem.* Let  $\phi(\mathbf{x})$  be a continuous point-to-point mapping of a closed simplex  $X$  into itself. Then there exists a point  $\mathbf{x}_0$  in

$X$  such that  $\mathbf{x}_0 = \phi(\mathbf{x}_0)$ . (For our purposes, we define a simplex in  $E^n$  as a convex set spanned by the origin and  $n$  linearly independent points.)

The theorem is actually correct for any set which is homeomorphic to an  $n$ -dimensional simplex.

The fixed-point theorem was generalized by Kakutani to cover point-to-set mappings. Prior to stating this result, we clarify a few terms. A mapping  $\phi$  which transforms every point  $\mathbf{x}$  in  $X$  into a subset of  $X$  is called a *point-to-set* mapping. A point-to-set mapping  $\phi$  is called *upper semicontinuous* if  $\mathbf{x}_n \rightarrow \mathbf{x}_0$ ,  $\mathbf{y}_n \rightarrow \mathbf{y}_0$ ,  $\mathbf{x}_n \in X$ ,  $\mathbf{y}_n \in \phi(\mathbf{x}_n)$  imply  $\mathbf{y}_0 \in \phi(\mathbf{x}_0)$ .

It is easily verified that a mapping  $\phi$  is upper semicontinuous if and only if the graph  $\{\langle \mathbf{x}, \mathbf{y} \rangle | \mathbf{x} \in X, \mathbf{y} \in \phi(\mathbf{x})\}$  is closed in  $E^{2n}$ .

A point-to-set mapping  $\phi(\mathbf{x})$  is said to be lower semicontinuous if for every  $\mathbf{y} \in \phi(\mathbf{x}_0)$  whenever  $\mathbf{x}_n \rightarrow \mathbf{x}_0$  there exists  $\mathbf{y}_n \in \phi(\mathbf{x}_n)$  such that  $\mathbf{y}_n \rightarrow \mathbf{y}$ .

A point-to-set mapping is continuous if it is both lower and upper semicontinuous. The definition of upper semicontinuous for a point-to-set mapping  $\phi$  reduces to that of Section C.1 when  $\phi$  is a point function, provided we define the containing relation  $\mathbf{y} \in \phi(\mathbf{x})$  as  $\mathbf{y} \leq \phi(\mathbf{x})$ . It is also possible to express the properties of point-to-set mappings in terms of neighborhood concepts. For example,  $\phi(\mathbf{x})$  is upper semicontinuous at  $\mathbf{x}_0$  if corresponding to any open set  $U$  containing  $\phi(\mathbf{x}_0)$  there exists some  $\delta(U) > 0$  such that  $d(\mathbf{x}, \mathbf{x}_0) < \delta$  implies  $\phi(\mathbf{x}) \subset U$ .

*Kakutani's fixed-point theorem.* Let  $X$  be a closed simplex, and let  $\phi$  represent an upper semicontinuous mapping which maps each point of  $X$  into a closed convex subset of  $X$ . Then there exists a point  $\mathbf{x}_0 \in X$  such that  $\mathbf{x}_0 \in \phi(\mathbf{x}_0)$ .

This fixed-point theorem has been further generalized by Eilenberg and Montgomery [73] and Begle [18].

**C.3 Set functions and probability distributions.** We review very briefly, without proofs, some of the theory behind the Lebesgue-Stieltjes integral. For full discussions, see Graves [112].

*The class of Borel sets.* The set of all points that belong to some member of a sequence  $\{S_n\}$  is the union of the sequence, and the set of all points that belong to every member of the sequence is called the intersection. These two sets are denoted by

$$\bigcup_{n=1}^{\infty} S_n \quad \text{and} \quad \bigcap_{n=1}^{\infty} S_n,$$

respectively. The set of points that belong to  $S_1$  but do not belong to  $S_2$  is called the difference, and it is denoted by  $S_1 - S_2$ .

A class  $\mathcal{S}$  of subsets of  $E^k$  is said to be an additive class of sets if it satisfies the following conditions:

- (1)  $E^k$  belongs to  $\mathcal{S}$ .  
 (2) If  $S_1, \dots, S_n, \dots$  all belong to  $\mathcal{S}$ , then

$$\bigcup_{n=1}^{\infty} S_n$$

belongs to  $\mathcal{S}$ .

- (3) If  $S_1$  and  $S_2$  belong to  $\mathcal{S}$ , then  $S_1 - S_2$  belongs to  $\mathcal{S}$ .

The smallest additive class of sets that contains all the “rectangles” in  $E^k$  (or all the intervals if  $k = 1$ ) is called the class of Borel sets of  $E^k$ .

*Nonnegative set functions.* A nonnegative set function  $P$  is a real-valued function that is defined on a class of Borel sets  $\mathcal{S}$  and satisfies the following conditions:

(1)  $P(S) \geq 0$  for all  $S \in \mathcal{S}$ .

(2)  $P\left(\bigcup_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} P(S_n)$  if  $S_n \cap S_m = \emptyset$  for  $n \neq m$ .

(3)  $P(S) < \infty$  if  $S$  is bounded.

The class of all nonnegative set functions defined on the class of Borel sets of real numbers is equivalent to the class of all monotone-increasing functions (defined on the real line) that are continuous from the right; this equivalence is unique if we can identify two increasing functions that differ everywhere by a fixed constant. The equivalence is constructed as follows: Given  $P$  and any real number  $\alpha$ , define

$$F(x, \alpha) = \begin{cases} P\{\xi | \alpha < \xi \leq x\} & x > \alpha, \\ 0 & x = \alpha, \\ -P\{\xi | x < \xi \leq \alpha\} & x < \alpha. \end{cases}$$

$F(x, \alpha)$  is an increasing function for each  $\alpha$ , and it is continuous from the right; furthermore, if  $\alpha_1 < \alpha_2$ , then

$$F(x, \alpha_1) - F(x, \alpha_2) = P\{\xi | \alpha_1 < \xi \leq \alpha_2\}.$$

That is,  $F(x, \alpha_1)$  and  $F(x, \alpha_2)$  differ by a constant, independent of  $x$ . On the other hand, if  $F(x)$  is any monotone-increasing function that is also continuous from the right, one may define

$$P\{\xi | a < \xi \leq b\} = F(b) - F(a).$$

The resulting set function defined on intervals can be extended to a non-negative set function defined on all Borel sets.

If  $P\{\xi|-\infty < \xi < \infty\} = 1$ , the set function is called a probability measure, and the corresponding point function  $F$  is commonly normalized so that  $F(-\infty) = 0$  and  $F(\infty) = 1$ . The function  $F$ , when normalized, is referred to as the (cumulative) distribution function. If  $F'(x) = f(x)$  exists, then  $f(x)$  is called the probability density function. Clearly,

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1.$$

Any function  $f$  that satisfies the last two conditions is the density function of some probability distribution.

These concepts can be extended to  $E^n$ . The probability distributions are the nonnegative set functions  $P$  with the property that  $P(E^n) = 1$ . To each  $P$  there corresponds a cumulative distribution function  $F$  defined by

$$F(x_1, \dots, x_n) = P\{\xi \in E^n | \xi_1 \leq x_1, \dots, \xi_n \leq x_n\},$$

etc. This function  $F$  is monotone-increasing in each variable;

$$F(-\infty, x_2, \dots, x_n) = \dots = F(x_1, x_2, \dots, x_{n-1}, -\infty) = 0$$

and

$$F(\infty, \infty, \dots, \infty) = 1.$$

Furthermore,  $F$  is continuous from the right in each variable; and the difference  $\Delta_1 \Delta_2 \dots \Delta_n F$ , where by definition

$$\begin{aligned} \Delta_i F &= F(a_1, \dots, a_{i-1}, a_i + \delta_i, a_{i+1}, \dots, a_n) \\ &\quad - F(a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n), \end{aligned}$$

is nonnegative since this difference is

$$P\{\xi | a_1 < \xi_1 \leq a_1 + \delta_1, \dots, a_n < \xi_n \leq a_n + \delta_n\}.$$

Conversely, any function with the last four properties is the cumulative distribution function of some probability distribution.

*Marginal distributions.* If  $F(x_1, \dots, x_n)$  is a cumulative distribution function defined on  $E^n$ , then  $F(x_1, \dots, x_{n-1}, \infty)$  is also a distribution function defined on  $E^{n-1}$ . In general, if  $k$  of the variables are assigned the value  $+\infty$ , then  $F$  induces a distribution function of the remaining  $n - k$  variables, which is called the marginal distribution.

*Lebesgue-Stieltjes integral.* Let  $g$  be a bounded function defined on  $E^n$ , and let  $P$  be a probability distribution defined on the class of Borel sets of

$E^n$ . In order to define the upper and lower integrals of  $g$  with respect to  $P$ , we consider first any subdivision of  $E^n$  into disjoint Borel sets  $E_1, \dots, E_k$  and define

$$m_i = \inf_{x \in E_i} g(x), \quad M_i = \sup_{x \in E_i} g(x),$$

$$s = \sum_{i=1}^k m_i P(E_i), \quad S = \sum_{i=1}^k M_i P(E_i).$$

The numbers  $s$  and  $S$  are called the lower and upper Darboux sums of  $g$  for the given subdivision. By varying the subdivision one obtains different values for the Darboux sums; if a subdivision is a refinement of another, the lower Darboux sum increases and the upper Darboux sum decreases. The supremum of the lower Darboux sums is called the lower integral of  $g$  with respect to  $P$ ; the infimum of the upper Darboux sums is called the upper integral. If both integrals are equal, then the function  $g$  is said to be integrable with respect to  $P$ , and its integral is the common value of the upper and lower integrals; this value is usually denoted by

$$\int g(x) dP(x).$$

If  $F$  is the point function associated with the set function  $P$ , the previous notation is commonly replaced by

$$\int g(x) dF(x).$$

Besides the obvious additive properties of integrals, we record for reference the Lebesgue criterion of dominated convergence. Let  $|f_n| \leq g$  and let  $g$  possess an integral with respect to  $P$ . If  $f_n$  converges everywhere to  $f$  (actually a.e. with respect to  $P$  would suffice), then

$$\lim_{n \rightarrow \infty} \int f_n dP = \int f dP. \quad (\text{C.3.1})$$

If  $f_n$  forms a monotone-increasing sequence of integrable functions with no other restrictions and  $P$  is a positive measure, then the same conclusion holds, provided that the value infinity is allowed.

*Sequences of distributions. Weak\* convergence.* A sequence  $\{F_n\}$  of distribution functions converges weak\* to a distribution function  $F$  if  $F_n(x) \rightarrow F_0(x)$  at every point of continuity of  $F_0$ . Every sequence of distribution functions contains a subsequence  $F_{n_k}$  that converges weak\* to a function  $F$ , but  $F$  need not be a distribution function. However, if

$g$  is continuous  $g(\pm\infty) = 0$ , and  $F_{n_k} \rightarrow F$  (weak\*), then

$$\lim_{k \rightarrow \infty} \int g(x) dF_{n_k}(x) = \int g(x) dF(x). \quad (\text{C.3.2})$$

If, in addition,  $F$  is a distribution function, the convergence of (C.3.2) is valid for any  $g$  which is bounded and continuous.

These results are known as *Helly's convergence theorem*.

## BIBLIOGRAPHY

Each reference work is listed once under the name of the first author as the article or book appears in the literature. At the end of the bibliography, each co-author is cited for the articles or books not previously listed directly under his name. In this case, the relevant reference is indicated by the appropriate numbered article in the bibliography.

- [1] ALLAIS, M., *Traité d'économie pure*, Imprimerie Nationale, Paris, (1952), 852 pp.
- [2] ALLEN, R. G. D., *Mathematical Economics*, London: Macmillan, 1956.
- [3] ARROW, K. J., "Alternative proof of the substitution theory for Leontief models in the general case," Chap. 9 of [146].
- [4] ARROW, K. J., "An extension of the basic theorems of classical welfare economics," in [200], pp. 507-32.
- [5] ARROW, K. J., E. W. BARANKIN, and D. BLACKWELL, "Admissible points of convex sets," in [162], pp. 87-91.
- [6] ARROW, K. J., D. BLOCK, and L. HURWICZ, "On the stability of the competitive equilibrium, II," *Econometrica*, **27** (1959), 82-109.
- [7] ARROW, K. J., and G. DEBREU, "Existence of an equilibrium for a competitive economy," *Econometrica*, **22** (1954), 265-90.
- [8] ARROW, K. J., T. E. HARRIS, and J. MARSCHAK, "Optimal inventory policy," *Econometrica*, **19** (1951), 250-72.
- [9] ARROW, K. J., and L. HURWICZ, "Reduction of constrained maxima to saddle-point problems," in [201], V, 1-20.
- [10] ARROW, K. J., and L. HURWICZ, "The gradient method for concave programming," Chap. 6 of [12].
- [11] ARROW, K. J., and L. HURWICZ, "On the stability of the competitive equilibrium, I," *Econometrica*, **26** (1958), 522-552.
- [12] ARROW, K. J., L. HURWICZ, and H. UZAWA (eds.), *Studies in Linear and Non-Linear Programming*, Stanford, Calif.: Stanford University Press, 1958.
- [13] ARROW, K. J., and M. McMANUS, "A note on dynamic stability," *Econometrica*, **26** (1958), 448-54.
- [14] ARROW, K. J., and M. NERLOVE, "A note on expectation and stability," *Econometrica*, **26** (1958), 297-305.
- [15] ARROW, K. J., and S. KARLIN, "Price speculation under certainty," Chap. 13 of [12].
- [16] ARROW, K. J., S. KARLIN, and H. SCARF, *Studies in the Mathematical Theory of Inventory and Production*, Stanford, Calif.: Stanford University Press, 1958.
- [17] BALDWIN, R. R., W. E. CANTEY, H. MAISEL, and J. P. McDERMOTT, "The Optimum Strategy in Blackjack," *J. Amer. Stat. Assoc.*, **51** (1956), 429-39.
- [18] BEGLE, E. G., "A fixed point theorem," *Ann. Math.*, **51** (1950), 544-50.
- [19] BELLMAN, R., *Dynamic Programming*, Princeton, N. J.: Princeton University Press, 1957.

- [20] BELLMAN, R., "On games involving bluffing," *Rendiconti del Circolo Mathematico di Palermo*, Series 2, Vol. 1 (1952), 139–56.
- [21] BELLMAN, R., and M. A. GIRSHICK, "An extension of results on duels with two opponents, one bullet each, silent guns, equal accuracy," The RAND Corporation, D-403, 1949.
- [22] BELLMAN, R., I. GLICKSBERG, and O. GROSS, "On some variational problems occurring in the theory of dynamic programming," *Rendiconti del Circolo Mathematico di Palermo*, Series 2, Vol. 3 (1954), 1–35.
- [23] BELLMAN, R., and M. SHIFFMAN, "On the Min-Max of
- $$\int_0^1 f(x)a(x) d(x) dt(x),$$
- The RAND Corporation, RM-308-1, 1949.
- [24] BELZER, R. L., "Silent duels, specified accuracies, one bullet each," The RAND Corporation, RAD(L)-301, 1948.
- [25] BERGE, C., "Sur une théorie ensembliste des jeux alternatifs," *Journal de mathématiques pures et appliquées*, 32 (1953), 129–84.
- [26] BIRKOFF, G., and S. MACLANE, *A Survey of Modern Algebra*, New York: Macmillan, 1941.
- [27] BLACKETT, D. W., "Some blotto games," *Naval Research Logistics Quarterly*, 1 (1954), 55–60.
- [28] BLACKWELL, D., "The silent duel, one bullet each, arbitrary accuracy," The RAND Corporation, RM-302, 1948.
- [29] BLACKWELL, D., "The noisy duel, one bullet each, arbitrary, non-monotone accuracy," The RAND Corporation, D-442, 1949.
- [30] BLACKWELL, D., and M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, New York: Wiley, 1954.
- [31] BLACKWELL, D., and M. A. GIRSHICK, "A loud duel with equal accuracy where each duelist has only a probability of possessing a bullet," The RAND Corporation, RM-219, 1949.
- [32] BLACKWELL, D., and M. SHIFFMAN, "A bomber-fighter duel," The RAND Corporation, D-509, 1949.
- [33] BLACKWELL, D., and M. SHIFFMAN, "A bomber-fighter duel," The RAND Corporation, RM-193, 1949.
- [34] BOHNENBLUST, H. F., "The Theory of Games," E. F. BECKENBACH (ed.), *Modern Mathematics for the Engineer*, New York: McGraw-Hill, 1956.
- [35] BOHNENBLUST, H. F., and S. KARLIN, "On a theorem of Ville," in [161], pp. 155–61.
- [36] BOHNENBLUST, H. F., S. KARLIN, and L. S. SHAPLEY, "Solutions of discrete two-person games," in [161], pp. 51–73.
- [37] BOHNENBLUST, H. F., S. KARLIN, and L. S. SHAPLEY, "Games with continuous, convex pay-off," in [161], pp. 181–92.
- [38] BOREL, E., "Applications aux jeux de hasard," *Traité du calcul des probabilités et de ses applications*, Paris: Gauthier-Villars, 1938.
- [39] BOREL, E., "The theory of play and integral equations with skew symmetrical kernels," "On games that involve chance and the skill of the players,"

- and "On systems of linear forms of skew symmetric determinants and the general theory of play," trans. L. J. Savage, *Econometrica*, **21** (1953), 97-117.
- [40] BROWN, G. W., "Some notes on computation of games solutions," The RAND Corporation, D-436, 1949.
- [41] BROWN, G. W., "Iterative solution of games by fictitious play," Chap. 24 of [146].
- [42] BROWN, G. W., and J. VON NEUMANN, "Solutions of games by differential equations," in [161], pp. 73-79.
- [43] BROWN, R. H., "The solution of a certain two-person zero-sum game," *Operations Research*, **5** (1957), 63-67.
- [44] CHARNES, A., "Optimality and degeneracy in linear programming," *Econometrica*, **20** (1952), 160-70.
- [45] CHARNES, A., and W. W. COOPER, "Generalizations of the warehousing model," *Operations Research Quarterly*, **6** (1955), 131-72.
- [46] CHARNES, A., W. W. COOPER, and A. HENDERSON, *Introduction to Linear Programming*, New York: Wiley, 1953.
- [47] CHENERY, H. B., "The role of industrialization in development programs," *Amer. Econ. Rev.*, **45** (1955), 40-56.
- [48] CHENERY, H. B., and K. S. KRETCHMER, "Resource allocation for economic development," *Econometrica*, **24** (1956), 365-99.
- [49] CHENERY, H. B., and H. UZAWA, "Non-linear programming in economic development," Chap. 15 of [12].
- [50] CHURCHMAN, C. W., R. L. ACKOFF, and E. L. ARNOFF, *Introduction to Operations Research*, New York: Wiley, 1957.
- [51] DANSKIN, J. M., "Fictitious play for continuous games," *Naval Research Logistics Quarterly*, **1** (1954), 313-20.
- [52] DANSKIN, J. M., "Mathematical treatment of a stockpiling problem," *Naval Research Logistics Quarterly*, **2** (1955), 99-109.
- [53] DANSKIN, J. M., and L. GILLMAN, "Explicit solution of a game over function space," The RAND Corporation, P-235, 1951.
- [54] DANSKIN, J. M., and L. GILLMAN, "A game over function space," *Rivista Mat. Univ. Parma*, **4** (1953), 83-94.
- [55] DANTZIG, G. B., "A proof of equivalence of the programming problem and the game problem," Chap. 20 of [146].
- [56] DANTZIG, G. B., "Maximization of a linear function of variables subject to linear inequalities," Chap. 21 of [146].
- [57] DANTZIG, G. B., *The Theory of Mathematical Programming*, notes, The RAND Corporation.
- [58] DANTZIG, G. B., and A. J. HOFFMAN, "Dilworth's theorem on partially ordered sets," in [163], pp. 207-14.
- [59] DANTZIG, G. B., A. ORDEN, and P. WOLFE, "The generalized simplex method for minimizing a linear form under linear inequality restraints," *Pac. J. Math.*, **5** (1955), 183-95.
- [60] DANTZIG, G. B., and A. WALD, "On the fundamental lemma of Neyman and Pearson," *Ann. Math. Stat.*, **22** (1951), 87-93.
- [61] DEBREU, G., "A social equilibrium existence theorem," *Proc. Nat. Acad. Sci.*, **38** (1952), 886-93.

- [62] DEBREU, G., "Valuation equilibrium and Pareto optimum," *Proc. Nat. Acad. Sci.*, **40** (1954), 588-92.
- [63] DEBREU, G., and I. N. HERSTEIN, "Non-negative square matrices," *Econometrica*, **21** (1953), 597-607.
- [64] DORFMAN, R., *Application of Linear Programming to the Theory of the Firm*, Berkeley and Los Angeles: University of California Press, 1951.
- [65] DORFMAN, R., P. A. SAMUELSON, and R. M. SOLOW, *Linear Programming and Economic Analysis*, New York: McGraw-Hill, 1958.
- [66] DRESHER, M., *Theory and Applications of Games of Strategy*, The RAND Corporation, R-216 (1951).
- [67] DRESHER, M., "Methods of solution in game theory," *Econometrica*, **18** (1950), 179-81.
- [68] DRESHER, M., "Games of strategy," *Mathematics Magazine*, **25** (1951), 93-99.
- [69] DRESHER, M., "Le Her," RAND report (unpublished).
- [70] DRESHER, M., and S. KARLIN, "Solutions of convex games as fixed points," in [162], pp. 75-86.
- [71] DRESHER, M., S. KARLIN, and L. S. SHAPLEY, "Polynomial games," in [161], pp. 161-180.
- [72] DRESHER, M., A. W. TUCKER, and P. WOLFE (eds.), *Contributions to the theory of Games, III*, (Ann. Math. Studies, Vol. 39), Princeton, N. J.: Princeton University Press, 1957.
- [73] EILENBERG, S., and D. MONTGOMERY, "Fixed point theorems for multi-valued transformations," *Amer. J. Math.*, **68** (1946), 214-22.
- [74] EISEMANN, K., "Linear programming," *Quart. Appl. Math.*, **13** (1955), 209-32.
- [75] ENTHOVEN, A. C., and K. J. ARROW, "A theorem on expectations and the stability of equilibrium," *Econometrica*, **24** (1956), 288-93.
- [76] FAN, K., "Fixed point and minimax theorems in locally convex topological linear spaces," *Proc. Nat. Acad. Sci. U. S. A.*, **38** (1952), 121-26.
- [77] FAN, K., "Minimax theorems," *Proc. Nat. Acad. Sci. U. S. A.*, **39** (1953), 42-47.
- [78] FAN, K., I. GLICKSBERG, and A. J. HOFFMAN, "Systems of inequalities involving convex functions," *Proc. Amer. Math. Soc.*, **8** (1957), 617-22.
- [79] FARKAS, J., "Theorie der einfachen Ungleichungen," *Journal für reine und angewandte Mathematik*, **124** (1901), 1-27.
- [80] FELLER, W., *An Introduction to Probability Theory and its Applications, I*, New York: Wiley, 1950.
- [81] FENCHEL, W., "Convex Cones, Sets, and Functions," Lecture notes, Department of Mathematics, Princeton University, 1953.
- [82] FERGUSON, R. O., and L. F. SARGENT, *Linear Programming*, New York: McGraw-Hill, 1958.
- [83] FLEMING, W. H., "On a class of games over function space and related variation problems," *Ann. Math.*, **60** (1954), 578-94.
- [84] FLOOD, M. M., "On the Hitchcock distribution problem," *Pac. J. Math.*, **3** (1953), 369-86.

- [85] FORD, L. R., JR., and D. R. FULKERSON, "Maximal flow through a network," *Canadian J. Math.*, **8** (1956), 399-404.
- [86] FORD, L. R., JR., and D. R. FULKERSON, "A simple algorithm for finding maximal network flows and an application to the Hitchcock Problem," *Canadian J. Math.*, **9** (1957), 210-18.
- [87] FORD, L. R., JR., and D. R. FULKERSON, "A primal dual algorithm for the capacitated Hitchcock problem," *Naval Research Logistics Quarterly*, **4** (1957), 47-54.
- [88] FRANK, M., and P. WOLFE, "An algorithm for quadratic programming," *Naval Research Logistics Quarterly*, **3** (1956), 95-110.
- [89] FROBENIUS, G., "Über Matrizen aus positiven Elementen," *Sitzungsberichte*, **8** (1908), 471-76.
- [90] FULKERSON, D. R., "Hitchcock transportation problem," The RAND Corporation, P-890, 1956.
- [91] FULKERSON, D. R., "Notes on linear programming. Part XLV. A network-flow feasibility theorem and combinatorial applications," The RAND Corporation, RM-2159, 1958.
- [92] GADDUM, J. W., A. J. HOFFMAN, and D. SOKOLOWSKY, "On the solution of the caterer problem," *Naval Research Logistics Quarterly*, **1** (1954), 223-29.
- [93] GALE, D., "The law of supply and demand," *Mathematica Scandinavica*, **3** (1955), 155-69.
- [94] GALE, D., "The closed linear model of production," in [163], pp. 285-303.
- [95] GALE, D., "Information in games with finite resources," in [72], pp. 141-45.
- [96] GALE, D., Lecture notes, Brown University (1957).
- [97] GALE, D., and O. GROSS, "A note on polynomial and separable games," The RAND Corporation, P-1216, 1957.
- [98] GALE, D., H. W. KUHN, and A. W. TUCKER, "On symmetric games," in [161], pp. 81-87.
- [99] GALE, D., H. W. KUHN, and A. W. TUCKER, "Linear programming and the theory of games," Chap. 19 of [146].
- [100] GALE, D., and S. SHERMAN, "Solutions of finite two-person games," in [161], pp. 37-49.
- [101] GANTMACHER, F., *Theory of Matrices* (in Russian), Moscow, 1954.
- [102] GANTMACHER, F., and M. KREIN, *Oscillatory Matrices and Kernels and Small Vibrations of Mechanical Systems*, 2d ed. (in Russian), Moscow, 1950.
- [103] GASS, S. I., *Linear Programming, Methods and Applications*, New York: McGraw-Hill, 1958.
- [104] GEORGESCU-ROEGEN, N., "The aggregate linear production function and its application to von Neumann's economic model," Chap. 4 of [146].
- [105] GILLIES, D. B., J. P. MAYBERRY, and J. VON NEUMANN, "Two variants of poker," in [162] pp. 13-50.
- [106] GILLMAN, L., "Operations analysis and the theory of games: an advertising example," *J. Amer. Stat. Assoc.*, **45** (1950), 541-46.
- [107] GLICKSBERG, I., "Noisy duel, one bullet each, with simultaneous fire and unequal worths," The RAND Corporation, RM-474, 1950.

- [108] GLICKSBERG, I., "A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points," *Proc. Amer. Math. Soc.*, **3** (1952), 170-74.
- [109] GLICKSBERG, I., and O. GROSS, "Butterfly games," The RAND Corporation, RM-655, 1951.
- [110] GLICKSBERG, I., and O. GROSS, "Notes on games over the square," in [162], pp. 173-84.
- [111] GOLDMAN, A. J., and A. W. TUCKER, "Theory of linear programming," in [163], pp. 53-97.
- [112] GRAVES, L. M., *The Theory of Functions of Real Variables*, New York: McGraw-Hill, 1946.
- [113] GROSS, O., "A rational payoff characterization of the Cantor distribution," The RAND Corporation, D-1349, 1952.
- [114] GROSS, O., "The derivatives of the value of a game," The RAND Corporation, RM-1286, 1954.
- [115] GROSS, O., "A simple linear programming problem explicitly solvable in integers," The RAND Corporation, RM-1560, 1955.
- [116] GROSS, O., "A rational game on the square," in [72], pp. 307-11.
- [117] HAHN, F. H., "Gross substitutes and the dynamic stability of general equilibrium," *Econometrica*, **26** (1958), 169-70.
- [118] HARDY, G. H., J. E. LITTLEWOOD, and G. PÓLYA, *Inequalities*, Cambridge, England: Cambridge University Press, 1952.
- [119] HAWKINS, D., and H. A. SIMON, "Note: Some conditions of macroeconomic stability," *Econometrica*, **17** (1949), 245-48.
- [120] HELMER, O., "Open problems in game theory," *Econometrica*, **20** (1952), 90 abstract.
- [121] HICKS, J. R., *Value and Capital*, 2d ed., Oxford: Oxford University Press, 1953.
- [122] HILDRETH, C., "A quadratic programming procedure," *Naval Research Logistics Quarterly*, **4** (1957), 79-85.
- [123] HIRSCH, W., and G. B. DANTZIG, "The fixed charge problem," The RAND Corporation, RM-1383, 1954.
- [124] HITCHCOCK, F. L., "Distribution of a product from several sources to numerous localities," *J. Math. Phys.*, **20** (1941), 224-30.
- [125] HOHN, F. E., and F. MODIGLIANI, "Production planning over time and the nature of the expectation and planning horizon," *Econometrica*, **23** (1955), 46-66.
- [126] HOUTHAKKER, H. S., "Revealed preference and the utility function," *Economica*, **17** (1950), 159-74.
- [127] HOUTHAKKER, H. S., "La forme des courbes d'Engel," *Cahiers du Séminaire d'Économétrie*, (1953), 59-66.
- [128] HURWICZ, L., "Programming in linear spaces," Chap. 4 of [12].
- [129] JACOBS, W. W., "The caterer problem," *Naval Research Logistics Quarterly*, **1** (1954), 154-65.
- [130] KAKUTANI, S., "A generalization of Brouwer's fixed point theorem," *Duke Math. Journal*, **8** (1941), 457-58.

- [131] KANTOROVITCH, L., "On the transformation of masses," *Dokl. Akad. Nauk USSR*, **37** (1942), 199–201.
- [132] KAPLANSKY, I., "A contribution to von Neumann's theory of games," *Ann. Math.*, **46** (1945), 474–79.
- [133] KARLIN, S., "Operator treatment of minmax principle," in [161], pp. 133–54.
- [134] KARLIN, S., "Reduction of certain classes of games to integral equations," in [162], pp. 125–58.
- [135] KARLIN, S., "The theory of infinite games," *Ann. Math.*, **58** (1953), 371–401.
- [136] KARLIN, S., "Polya type distributions, II," *Ann. Math. Stat.*, **28** (1957), 281–308.
- [137] KARLIN, S., "On games described by bell-shaped kernels," in [72], pp. 365–91.
- [138] KARLIN, S., and R. RESTREPO, "Multistage poker models," in [72], pp. 337–63.
- [139] KARLIN, S., "Positive Operators," *Journal of Math. and Mech.*, (1960).
- [140] KARLIN, S., and L. S. SHAPLEY, "Some applications of a theorem on convex functions," *Ann. Math.*, **52** (1950), 148–53.
- [141] KARLIN, S., and L. S. SHAPLEY, "Geometry of Moment Spaces," *Memoirs of the American Mathematics Society*, **12** (1953), 1–93.
- [142] KEMENY, J. G., O. MORGENSEN, and G. L. THOMPSON, "A generalization of the von Neumann model of an expanding economy," *Econometrica*, **24** (1956), 115–35.
- [143] KIEFER, J., "Invariance, minimax sequential estimation and continuous time processes," *Ann. Math. Stat.*, **28** (1957), 573–601.
- [144] KNIGHT, F. H., "A note on Professor Clark's illustration of marginal productivity," *Journal of Political Economics*, **33** (1925), 550–53.
- [145] KOOPMAN, B. O., "The theory of search, I: kinematic bases," *Operations Research*, **4** (1956), 324–46.
- [146] KOOPMANS, T. C. (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13, New York: Wiley, 1951.
- [147] KOOPMANS, T. C., *Three Essays on the State of Economic Analysis*, New York: McGraw-Hill, 1957.
- [148] KOOPMANS, T. C., "Alternative proof of the substitution theorem for Leontief models in the case of three industries," Chap. 8 of [146].
- [149] KOOPMANS, T. C., "Water storage policy in a simplified hydroelectric system," in *Proc. International Conference on Operations Research*, Bristol, England: John Wright and Sons, 1958.
- [150] KOOPMANS, T. C., and S. REITER, "A model of transportation," Chap 14 of [146].
- [151] KOTELYANSKII, D. M., "On some properties of matrices with positive elements," (in Russian), *Mat. Sbornik (N. S.)*, **31** (1952), 497–506.
- [152] KOTELYANSKII, D. M., "On a property of sign-symmetric matrices," *Uspehi Matem. Nauk (N. S.)*, **8** (1953), 163–67.
- [153] KREIN, M. G., and M. A. RUTMAN, "Linear operators leaving invariant a

- cone in a Banach space," *Uspehi Matem. Nauk (N. S.)*, **3** (1948), 3–95 (Amer. Math. Soc. Translation No. 26, New York, 1950).
- [154] KRETCHMER, K. S., "Linear programming in locally convex spaces and its use in analysis," Ph.D. thesis, Carnegie Inst. of Tech.
- [155] KUDO, H., "On minimax invariant estimates of the transformation parameter," *National Science Report of the Ochanomizu University*, **6** (1955), 31–73.
- [156] KUHN, H. W., "A simplified two-person poker," in [161], pp. 97–103.
- [157] KUHN, H. W., "A combinatorial algorithm for the assignment problem," Issue 11, Logistics Papers, George Washington University Logistics Research Project, 1954.
- [158] KUHN, H. W., "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, **2** (1955), 83–97.
- [159] KUHN, H. W., "On a theorem of Wald," in [163], pp. 265–73.
- [160] KUHN, H. W., "Variants of the Hungarian method for assignment problems," *Naval Research Logistics Quarterly*, **3** (1956), 253–58.
- [161] KUHN, H. W., and A. W. TUCKER (eds.), *Contributions to the Theory of Games, I*, (Ann. Math. Studies, Vol. 24), Princeton, N. J.: Princeton University Press, 1950.
- [162] KUHN, H. W., and A. W. TUCKER (eds.), *Contributions to the Theory of Games, II*, (Ann. Math. Studies, Vol. 28), Princeton, N. J.: Princeton University Press, 1953.
- [163] KUHN, H. W., and A. W. TUCKER (eds.), *Linear Inequalities and Related Systems*, (Ann. Math. Studies, Vol. 38), Princeton, N. J.: Princeton University Press, 1956.
- [164] KUHN, H. W., and A. W. TUCKER, "Non-linear programming," in [200], pp. 481–92.
- [165] LEMKE, C. E., "The dual method of solving the linear programming problem," *Naval Research Logistics Quarterly*, **1** (1954), 36–47.
- [166] LEONTIEF, W. W., *The Structure of the American Economy*, Cambridge, Mass: Harvard University Press, 1941.
- [167] *Linear Programming*, Vols. I and II, Second Symposium, National Bureau of Standards, U. S. Department of Commerce, January 27–29, 1955.
- [168] LOOMIS, L. H., "On a theorem of von Neumann," *Proc. Nat. Acad. Sci. U. S. A.*, **32** (1946), 213–15.
- [169] LUCE, R. D., and H. Raiffa, *Games and Decisions*, New York: Wiley, 1957.
- [170] LUCE, R. D., and A. W. TUCKER (eds.), *Contributions to the Theory of Games, IV* (Ann. Math. Studies, Vol. 40), Princeton, N. J.: Princeton University Press, 1959.
- [171] MANNE, A. S., *Scheduling of Petroleum Refinery Operations*, Cambridge, Mass.: Harvard University Press, 1956.
- [172] MARKOWITZ, H. M., "Portfolio selection," *Journal of Finance*, **7** (1952), 77–91.
- [173] MARKOWITZ, H. M., "The optimization of a quadratic function subject to linear constraints," *Naval Research Logistics Quarterly*, **3** (1956), 111–13.
- [174] MARKOWITZ, H. M., and A. S. MANNE, "On the solution of discrete programming problems," *Econometrica*, **25** (1957), 84–110.

- [175] McCLOSKEY, J. F., and F. N. TREFETHEN, *Operations Research for Management*, Baltimore: Johns Hopkins University Press, 1954.
- [176] McDONALD, J., *Strategy in Poker, Business and War*, New York: Norton, 1950.
- [177] MCKENZIE, L. W., "Competitive equilibrium with dependent consumer preferences," in [167], 277-94.
- [178] MCKENZIE, L. W., "Demand theory without a utility index," *Review of Economic Studies*, **24** (1956-57), 185-89.
- [179] MCKINSEY, J. C. C., *Introduction to the Theory of Games*, New York: McGraw-Hill, 1952.
- [180] METZLER, L. A., "Stability of multiple markets: the Hicks conditions," *Econometrica*, **13** (1945), 277-92.
- [181] MILLS, H. D., "Marginal values of matrix games and linear programs," in [163], pp. 183-93.
- [182] MORGENSTERN, O. (ed.), *Economic Activity Analysis*, New York: Wiley, 1954.
- [183] MORISHIMA, M., "On the laws of change of the price system in an economy which contains complementary commodities," *Osaka Economics Papers*, **1** (1952), 101-13.
- [184] MORISHIMA, M., "Prices, interest and profits in a dynamic Leontief system," *Econometrica*, **26** (1958), 358-80.
- [185] MOSAK, J. L., *General Equilibrium Theory in International Trade*, Bloomington, Indiana: The Principia Press, 1944.
- [186] MOTZKIN, T. S., *Beiträge zur Theorie der linearen Ungleichungen*, [Inaugural Dissertation] Jerusalem, 1936.
- [187] MOTZKIN, T. S., H. RAIFFA, G. L. THOMPSON, and R. M. THRALL, "The double description method," in [162], pp. 51-73.
- [188] NASH, J. F., "Equilibrium points in  $N$ -person games," *Proc. Nat. Acad. Sci., U. S. A.*, **36** (1950), 48-49.
- [189] NASH, J. F., and L. S. SHAPLEY, "A simple three-person poker game," in [161], pp. 105-16.
- [190] NEYMAN, J., and E. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society, A-231* (1933), 289-337.
- [191] NIKAIDO, H., "On a minimax theorem and its applications to functional analysis," *Journal of the Mathematical Society of Japan*, **5** (1953), 86-94.
- [192] NIKAIDO, H., "Note on the general economic equilibrium for non-linear production functions," *Econometrica*, **22** (1954), 49-53.
- [193] NIKAIDO, H., "On von Neumann's minimax theorem," *Pac. J. Math.*, **4** (1954), 65-72.
- [194] PEISAKOFF, M. P., "Transformation parameters," Ph.D. thesis, Princeton University, 1950.
- [195] PERLIS, S., *Theory of Matrices*, Cambridge, Mass.: Addison-Wesley, 1952.
- [196] PITMAN, E. J. G., "The estimation of location and scale parameters of a continuous population of any given form," *Biometrika*, **30** (1939), 391-421.
- [197] PITMAN, E. J. G., "Tests of hypotheses concerning location and scale parameters," *Biometrika*, **31** (1939), 200-215.

- [198] PRAGER, W., "On the role of congestion in transportation problems," *Zeit. für angewandte Math. Mech.*, **35** (1955), 381-96.
- [199] PRAGER, W., "On the caterer problem," *Management Science*, **3** (1956), 15-23.
- [200] *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles: University of California Press, 1951.
- [201] *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles: University of California Press, 1955. 5 vols.
- [202] RADEMACHER, H., and I. J. SCHOENBERG, "Helly's theorems on convex domains and Tchebycheff's approximation problem," *Canadian Math.*, **2** (1950), 245-56.
- [203] REINFELD, N. V., and W. R. VOGEL, *Mathematical Programming*, Englewood Cliffs, N. J.: Prentice-Hall, 1958.
- [204] REITER, S., "Efficiency and prices in the theory of an international economy," Technical Report No. 13, Stanford University, 1954.
- [205] RESTREPO, R., "Tactical problems involving several actions," in [72], pp. 313-35.
- [206] RESTREPO, R., "Games with a random move," unpublished.
- [207] RILEY, V., and S. I. GASS, *A Bibliography of Linear Programming and Associated Techniques*, Baltimore: Johns Hopkins University Press, 1958.
- [208] ROBINSON, J., "An iterative method of solving a game," *Ann. Math.*, **54** (1951), 296-301.
- [209] RUDIN, W., *Principles of Mathematical Analysis*, New York: McGraw-Hill, 1953.
- [210] SAMUELSON, P. A., *Foundations of Economic Analysis*, Cambridge, Mass.: Harvard University Press, 1955.
- [211] SAMUELSON, P. A., "A note on the pure theory of consumers' behavior," *Economica*, **18** (1938), 61-71, 353-54.
- [212] SAMUELSON, P. A., "Frank Knight's theorem in linear programming," The RAND Corporation, D-782, 1950 (holographed).
- [213] SAMUELSON, P. A., "Abstract of a theorem concerning substitutability in open Leontief models," Chap. 7 of [146].
- [214] SAVAGE, I. R., "Cycling," *Naval Research Logistics Quarterly*, **3** (1956), 163-75.
- [215] SCHOENBERG, I. J., "On smoothing operations and their generating functions," *Bull. Amer. Math. Soc.*, **59** (1953), 199-230.
- [216] SHAPLEY, L. S., *The Theory of n-Person Games*, unpublished notes.
- [217] SHAPLEY, L. S., "The silent duel, one bullet *vs.* two, equal accuracy," The RAND Corporation, RM-445, 1950.
- [218] SHAPLEY, L. S., "An example of an infinite non-constant-sum game," The RAND Corporation, RM-898, 1952.
- [219] SHAPLEY, L. S., and R. N. SNOW, "Basic solutions of discrete games," in [161], pp. 27-37.
- [220] SHERMAN, S., "Games and subgames," *Proc. Amer. Math. Soc.*, **2** (1951), 186-87.

- [221] SHIFFMAN, M., "On the equality of  $\min \max = \max \min$ , and the theory of games," The RAND Corporation, RM-243, 1941.
- [222] SHIFFMAN, M., "Games of timing," in [162], pp. 97-123.
- [223] SION, M., and P. WOLFE, "On a game without a value," in [72], pp. 299-305.
- [224] SION, M., "On general minimax theorems," *Pac. J. Math.*, **8** (1958), 171-76.
- [225] SLUTSKY, E., "Sulla teoria del bilancio del consumatore," *Giornale degli Economisti*, **51** (1915), 19-23.
- [226] SMITH, W. E., "Various optimizers for single-stage production," *Naval Research Logistics Quarterly*, **3** (1956), 59-66.
- [227] SOLOW, R. M., "On the structure of linear models," *Econometrica*, **20** (1952), 29-46.
- [228] SOLOW, R. M., and P. A. SAMUELSON, "Balanced growth under constant returns to scales," *Econometrica*, **21** (1953), 412-24.
- [229] STIGLER, G. J., "The cost of substance," *Journal of Farm Economics*, **27** (1945), 303-14.
- [230] SUITS, D. B., "Dynamic growth under diminishing returns to scales," *Econometrica*, **22** (1954), 496-501.
- [231] THOMPSON, G. L., "On the solution of a game-theoretic problem," in [163], pp. 275-84.
- [232] UZAWA, H., "On the logical relation between preference and revealed preference," Technical Report No. 38, Department of Economics, Stanford University, 1956.
- [233] UZAWA, H., "A note on the stability of equilibrium," Technical Report No. 44, Department of Economics, Stanford University, 1957.
- [234] UZAWA, H., "A note on the Menger-Wieser theory of imputation," Technical Report No. 45, Department of Economics, Stanford University, 1957.
- [235] UZAWA, H., "The gradient method for concave programming, II. Global results," Chap. 7 of [12].
- [236] VAJDA, S., *Theory of Games and Linear Programming*, New York: Wiley, 1956.
- [237] VILLE, J., "Note sur la théorie générale des jeux où intervient l'habileté des joueurs," in "Applications aux jeux de hasard," by E. Borel and J. Ville, Tome IV, Fascicule II, in Borel [1938].
- [238] VON NEUMANN, J., "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, **100** (1928), 295-320.
- [239] VON NEUMANN, J., "Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," *Ergebnisse eines Mathematischen Kolloquiums*, **8** (1937), 73-83.
- [240] VON NEUMANN, J., "A certain zero-sum two-person game equivalent to the optimum assignment problem," in [162], pp. 5-12.
- [241] VON NEUMANN, J., "A numerical method to determine optimum strategy," *Naval Research Logistics Quarterly*, **1** (1954), 109-15.
- [242] VON NEUMANN, J., and O. MORGENTERN, *Theory of Games and Economic Behavior*, 2d ed., Princeton, N. J.: Princeton University Press, 1947 (1st ed., 1944).

- [243] WAGNER, H., "A practical guide to the dual theorem," *Operations Research*, **6** (1958), 364-84.
- [244] WALD, A., *Sequential Analysis*, New York: Wiley, 1947.
- [245] WALD, A., *Statistical Decision Functions*, New York: Wiley, 1950.
- [246] WALD, A., "Über die eindeutige positive Lösbarkeit der neuen Produktionsgleichungen," *Ergebnisse eines mathematischen Kolloquiums*, **6** (1935), 12-20.
- [247] WALD, A., "Über die Produktionsgleichungen der ökonomischen Wertlehre, II," *Ergebnisse eines mathematischen Kolloquiums*, **7** (1936), 1-6.
- [248] WALD, A., "Statistical decision functions which minimize the maximum risk," *Ann. Math.*, **46** (1945), 265-80.
- [249] WALD, A., "Statistical decision theory," *Ann. Math. Stat.*, **20** (1949), 165-205.
- [250] WALD, A., "Note on zero-sum two-person games," *Ann. Math.*, **52** (1950), 739-42.
- [251] WALD, A., "On some systems of equations of mathematical economics," *Econometrica*, **19** (1951), 368-403.
- [252] WALRAS, L., *Elements of Pure Economics, or the Theory of Social Wealth*, trans. W. Jaffe, Homewood, Ill.: Irwin, 1954.
- [253] WEYL, H., "The elementary theory of convex polyhedra," in [161], pp. 3-18.
- [254] WILLIAMS, J. D., *The Compleat Strategyst*, New York: McGraw-Hill, 1954.
- [255] WOLFE, P., "Determinateness of polyhedral games," in [163], pp. 195-98.
- [256] WOLFE, P., "The simplex method for quadratic programming," The RAND Corporation, P-1295, 1957.
- [257] WOOD, M. K., and G. B. DANTZIG, "Programming of interdependent activities, I. General discussion," *Econometrica*, **17** (1949), 193-99. (Reprinted in revised form as Chap. 1 of [146].)

- ACKOFF, R. L., see [50]
- ARNOFF, E. L., see [50]
- ARROW, K. J., see [75]
- BARANKIN, E. W., see [5]
- BLACKWELL, D., see [5]
- BLOCK, D., see [5]
- CANTEY, W. E., see [17]
- COOPER, W. W., see [45], [46]
- DANTZIG, G. B., see [123], [257]
- DEBREU, G., see [7]
- FULKERSON, D. R., see [85], [86], [87]
- GASS, S. I., see [207]
- GILLMAN, L., see [53], [54]
- GIRSCHICK, M. A., see [21], [30], [31], [78]
- GLICKSBERG, I., see [22], [78]
- GROSS, O., see [22], [97], [109], [110]
- HARRIS, T. E., see [8]
- HENDERSON, A., see [46]
- HOFFMAN, A. J., see [58], [78], [92]
- HURWICZ, L., see [6], [9], [10], [11], [12]
- KARLIN, S., see [15], [16], [35], [36], [37], [70], [71]
- KREIN, M., see [102]
- KRETCHMER, K. S., see [48]
- KUHN, H. W., see [98], [99]
- LITTLEWOOD, J. E., see [118]
- MACLANE, S., see [26]
- MAISEL, H., see [17]
- MANNE, A. S., see [174]
- MARSCHAK, J., see [8]
- MAYBERRY, J. P., see [105]
- McDERMOTT, J. P., see [17]
- McMANUS, M., see [13]
- MODIGLIANI, F., see [125]
- MORGENSTERN, O., see [142], [242]

- MONTGOMERY, D., see [73]  
NERLOVE, M., see [14]  
ORDEN, A., see [59]  
PEARSON, E., see [190]  
POLYA, G., see [118]  
RAIFFA, H., see [169], [187]  
REITER, S., see [150]  
RESTREPO, R., see [138]  
RUTMAN, M. A., see [153]  
SAMUELSON, P. A., see [65], [228]  
SARGENT, L. G., see [82]  
SCARF, H., see [16]  
SCHOENBERG, I. J., see [202]  
SHAPLEY, L. S., see [36], [37], [71],  
[140], [141], [189]  
SHERMAN, S., see [100]  
SHIFFMAN, M., see [23], [32], [33]  
SIMON, H. A., see [119]  
SNOW, R. N., see [219]  
SOKOLOWSKY, D., see [92]  
SOLOW, R. M., see [65]  
THRALL, R. M., see [187]  
THOMPSON, G. L., see [142], [187]  
TREFETHEN, F. N., see [175]  
TUCKER, A. W., see [72], [98], [99],  
[111], [161], [162], [163], [164], [170]  
UZAWA, H., see [12], [49]  
VOGEL, W. R., see [203]  
VON NEUMANN, J., see [42], [105]  
WALD, A., see [60]  
WOLFE, P., see [59], [72], [88], [223]

## INDEX

- activity analysis problem, 116, 117, 258  
adjoint matrix, 49, 387  
advertising example, 92  
algebraic multiplicity, 247  
anticipated future prices, 329  
Arrow-Hurwicz gradient method, 212  
artificial variables, 167  
  
balanced growth, 336, 339  
    rate of, 339  
bargaining example, 94  
barycentric coordinates, 43  
basis, 363  
Blotto game, 82, 99  
Brouwer fixed-point theorem, 275, 327,  
    408  
Brown algorithm, 159  
bluffing, 92  
  
caterer's problem, 143, 147  
characteristic equation, 387  
cofactor, 385  
compensated income function for price  
     $p$ , 271  
competitive equilibrium, 281, 295  
    stability of, 301  
complementary commodities, 307, 333  
compound matrices, 394  
concave functions, 404  
conjugate functions, 218, 226  
    composition of, 229  
constant returns to scale, 338  
consumer choice (theory of), 265  
convex cone (notation), 11  
convex cones, 402  
    duality theory of, 120  
    polyhedral, 403  
convex functions, 404  
convex hull of sets, 222, 399  
convex linear, 17, 34  
convex set (notation), 11  
convex set  
    direct sums of, 222  
    polyhedral, 72, 114  
cyclic matrix, 54  
  
demand function, 268  
determinant, 10  
determinants, 385  
diet problem, 114  
direct sums, 363  
dominance principle, 38, 40  
dominated convergence, 412  
doubly stochastic matrices, 132  
dual constraints, 117  
dual polyhedral cone, 120  
dual problem, 169  
Duality Principle, 220, 221, 279  
Duality Theorem  
    of linear programming, 122, 126  
    of nonlinear programming, 227  
  
eigenvalues, 371  
eigenvectors, 371  
efficient production, 342  
efficient vector, 216, 218, 262  
elasticity of expectations, 333  
equalizer, 38  
equicontinuous functions, 408  
equilibrium  
    theory, 265  
    nonlinear models, 273  
equilibrium and exchange linear  
    models, 256  
essential strategy, 37  
Euclidean spaces, 362  
Euler relation, 309  
excess demand function, 303  
Existence Theorem (of linear pro-  
    gramming), 123  
expanding economy, 335  
    von Neumann Model, 340

- extreme points, 132, 161, 401  
 feasible vector, 114, 216, 262, 293  
 game  
 $2 \times m$ , 41  
 $3 \times m$ , 43  
 completely mixed, 51, 55, 264  
 essential part of, 65  
 I.F.F., 84, 99  
 ordered set, 53  
 reduced, 65, 69  
 geometrical multiplicity, 247  
 gradient differential equations, 199  
 gradient equations, 212  
 gross complements, 293  
 gross substitutes, 293, 305, 312  
 Helly's Theorem, 230, 413  
 hermitian matrices, 378  
 homogeneous functions, 326  
 inner product (definition), 10  
 interindustry nonlinear programming model, 206  
 Jordan canonical form, 371, 372  
 Kakutani fixed-point theorem, 28, 279, 409  
 kernel of extreme-point solutions, 48  
 Kuhn, Tucker Theorem, 200  
 Lagrange multiplier, 200  
 Lagrangian form, 118, 121, 259, 296  
 LeChatelier Principle, 153  
 Le Her Game, 101  
 Leontief model  
     closed, 244  
     dynamic, 338, 340  
     efficient points of, 261  
     open, 244  
     open and closed linear, 244  
 linear equations, 369  
 linear independence, 363  
 linear programming problem, 113  
 linear programming problem and its dual, 116  
 linear transformations, 366  
 local stability, 306  
 matrix decomposability, 254  
 matrix games (definition), 17  
 matrix, stable, 331  
 $M$ -matrix, 331, 332  
 maximal flow, 139  
 maximal flow, minimal-cut theorem, 140  
 Metzler matrices, 255  
 Minkowski-Leontief matrix, 52  
 min-max theorem, 24, 26, 28  
 network flow problems, 138, 176  
 Neyman-Pearson Theory, 9  
 nondegeneracy assumption, 161, 167  
 nonsingular matrices, 387  
 null space of a linear transformation, 68, 366  
 numéraire, 304  
 optimal strategies, assignment problem, 131  
     definition, 22  
     extreme-point, 36, 44  
 optimal strategies uniqueness, 70, 74  
 orthogonal complements, 68, 362, 363, 378  
 orthogonal projection, 380  
 Pareto optimum, 292, 294  
 pay-off kernel, 16  
 permutation matrix, 132  
 poker game, 20, 86, 99  
 polynomial approximation problem, 231  
 portfolio selection, 205  
 positive definite, 382  
 preference relation, 265  
 price adjustment processes, global stability, 310  
     normalized, 304, 308  
 price adjustment process, non-normalized, 304

- price speculation model, 148, 151  
production coefficients, 245  
production possibility sets, 260  
production theory, 258  
programming, concave, 200  
  nonlinear, 199  
  
quadratic form, 381  
quadratic programming algorithm,  
  206, 241  
quasi-definitie negative, 322, 332  
quasi-stable, 323, 324  
  
random variables, 79  
range of linear mapping, 68, 366  
rank, 68  
rates of transformations, nonincreasing  
  marginal, 338  
revealed preference, axiom of, 269  
  weak axiom, 311  
  
saddle point, 28, 110, 119, 124, 200,  
  213, 229, 259  
saddle-value problem, 259  
Samuelson Substitution Theorem, 261  
self-adjoint matrix, 247  
semicontinuous and equicontinuous  
  functions, 407  
shadow-price, 117  
similar transformation, 376  
simplex algorithm, dual, 17, 170  
  primal, 170  
  
simplex method, 159, 165, 174  
skew-symmetric matrix, 125  
“slack” variables, 160  
stability and expectations, 329  
static expectations, 330, 333  
statistical decision theory, 8  
strategies, geometrical representation  
  of, 26  
strategy, behavior, 21  
  definition, 15  
  mixed, 18  
  pure, 17  
  
tableau, 163  
tactical games of continuous firepower,  
  86  
transportation problem, 115, 117, 136  
transpose matrix, 10, 68, 378  
two-finger Morra, 20  
  
unitary spaces, 362  
  
value of a game (computation), 179  
Vandermonde determinant, 77  
vector maximum problem, 216  
vector system, 182  
  
Walras law, 302, 306  
warehouse storage problem, 115, 129  
weight (of a strategy), 64, 67, 74  
welfare economics, 293