A study an analysis of stochastic linear programming

Chamberlain Lambros Foes
Portland State University

5-1-1970

Recommended Citation

10.15760/etd.821

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
AN ABSTRACT OF THE THESIS OF Chamberlain Lambros Poes

Title: A Study and Analysis of Stochastic Linear Programming.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

This essay investigates the concept of linear programming in general and linear stochastic programming in particular. Linear stochastic programming is described as the model where the parameters of the linear programming admit random variability.

The first three chapters present through a set-geometric approach the foundations of linear programming. Chapter one describes the evolution of the concepts which resulted in the adoption of the model. Chapter two describes the constructs in n-dimensional euclidian space which constitute the
mathematical basis of linear programs, and chapter three defines the linear programming model and develops the computational basis of the simplex algorithm.

The second three chapters analyze the effect of the introduction of risk into the linear programming model. The different approaches of estimating and measuring risk are studied and the difficulties arising in formulating the stochastic problem and deriving the equivalent deterministic problems are treated from the theoretical and practical point of view. Multiple examples are given throughout the essay for clarification of the salient points.
A STUDY AND ANALYSIS OF
STOCHASTIC LINEAR PROGRAMMING

by

CHAMBERLAIN LAMBROS FCE3

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
APPLIED SCIENCE

Portland State University
1970
TO THE OFFICE OF GRADUATE STUDIES:

The members of the Committee approve the thesis of Chamberlain Lambros Foes presented May 22, 1970.

APPROVED:

Chairman

Head, Department of Applied Science

Dean of Graduate Studies
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
</tbody>
</table>

## CHAPTER

### I  THE CONCEPT OF LINEAR PROGRAMMING  

- Introduction ........................................... 1
- Optimization and Mathematical Programming ............... 5
- Current State of Mathematical Programming ................ 9
- Formulation of a Linear Programming Model ................ 11
- Characteristic Limitations of the Linear Programming Model ........................................... 14
- Some Linear Programming Examples ........................ 16

### II  MATHEMATICAL THEORY OF LINEAR SPACES  

- Introduction ........................................... 29
- Mathematical Constructs in N-Space General Definitions ........................................... 29
- Elements of the Theory of Matrices and Determinants ........................................... 43
- Linear Transformations ........................................... 57
III CONVEX SPACES AND LINEAR PROGRAMMING

Introduction
Basic Properties of Convex Sets
The General Linear Programming Problem
The Simplex Procedure

IV DECISION-MAKING UNDER RISK AND UNCERTAINTY

Introduction
Elements of Utility Theory
Characterization of Decision Process
Sources of Uncertainty in the Model Data
Errors and Stability in Linear Programming

V STOCHASTIC LINEAR PROGRAMMING

Introduction
Passive Stochastic Programming
Active Stochastic Programming

VI DUALITY IN STOCHASTIC PROGRAMS

Introduction
The Passive Case
The Active Case

VII CONCLUSIONS

BIBLIOGRAPHY
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>17</td>
</tr>
<tr>
<td>II</td>
<td>23</td>
</tr>
<tr>
<td>III</td>
<td>27</td>
</tr>
<tr>
<td>IV</td>
<td>38</td>
</tr>
<tr>
<td>V</td>
<td>117</td>
</tr>
</tbody>
</table>

I Atlas Manufacturing System .......................... 17
II Data of Feed Blend .................................... 23
III Input-Output Table .................................. 27
IV Annual Production in Tons ............................ 38
V Expected Values of Z .................................. 117
# TABLE OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Graphical solution of Atlas system</td>
</tr>
<tr>
<td>2</td>
<td>Graphical solution of Car Manufacturing system</td>
</tr>
<tr>
<td>3</td>
<td>Feed Blend Constraint Plot</td>
</tr>
<tr>
<td>4</td>
<td>Euclidian Plan</td>
</tr>
<tr>
<td>5</td>
<td>Linear Variety in S, $\mathbb{R}^3$</td>
</tr>
<tr>
<td>6</td>
<td>Parallelogram in $\mathbb{R}^2$</td>
</tr>
<tr>
<td>7</td>
<td>Parallelepiped in $\mathbb{R}^3$</td>
</tr>
<tr>
<td>8</td>
<td>Illustration of a Hyperplane in $\mathbb{R}^3$</td>
</tr>
<tr>
<td>9</td>
<td>Convex Sets in $\mathbb{R}^3$</td>
</tr>
<tr>
<td>10</td>
<td>Convex Polyhedron</td>
</tr>
</tbody>
</table>
CHAPTER I

THE CONCEPT OF LINEAR PROGRAMMING

On the gate is an inscription in every human tongue: "Mathematics spoken. Ici on parle mathematique. Hier spricht man mathematik"

INTRODUCTION

The requirements of physical problems spurned and shaped the development of mathematics at the end of the 19th and the beginning of the 20th century. In 1931, a paper was published in Hungary on a particular case of a transportation problem, the assignment problem. Professor Egervary who published the paper had not found an effective solution to the problem. In 1939, L. V. Kantorovich gave a report to the Leningrad State University on some mathematical methods of organizing and planning production. He presented a whole series of problems confronting the economic planners who wanted to achieve maximum production on the basis of the most efficient utilization of materials, labor and equipment.


He called his solution the method of resolving multipliers and he applied it for the solution of some of the problems he presented. Still, his method came short of being a universal method for the solution of problems of optimization under constraints. In 1942, Kantorovich published a small paper in terse mathematical language which laid down in brief the mathematical foundations of linear programming. ⁴ The interesting thing, however, was that at the time nobody recognized the significance of the paper. ⁵ But, even in this paper the problem of determining an effective method for solution of a specific problem was not solved.

During the Second World War governments found themselves managing a very complex war and wishing to increase the effectiveness of their weapons turned deliberately to science and the scientists for help. In England, a team of scientists nicknamed Blackett's Circus ⁶ undertook research into problems ranging from planning the destruction of enemy submarines and the safest routing of military convoys to the assessment of the flow of enemy munitions between war theaters. A new breed of scientists developed out of Blackett's Circus circles.


⁵Although the Kantorovich paper was published in Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS in volume XXXVII, No. 7-8, in 1942, it was not translated until sixteen years later

...the analysts. The modus operandi of these analysts when dealing with a set extremely complex in their interrelationships operations was to break down the problem into classes and then analyse it piece by piece. After the war, the analysts continued their work on peaceful and Cold War uses of their newly-found analytical capabilities.

In 1947, George Dantzig, working on a project for the U. S. Air Force, formulated the linear programming mathematical model which unified many of the seemingly diverse problems of the analysts. He also invented the first efficient general computational method, the simplex method, for explicit formulation of such problems and determination of their solutions. At about the same time, John Von Neumann, heading a group of mathematicians and engineers, delivered a course of lectures on "The Theory and Techniques of Electronic Digital Computers." The first electronic digital computer - the famous ENIAC has been completely in operation in the University of Pennsylvania since 1946. These two events were of cardinal importance for the development of the applications of linear programming: because, most of the linear programming problems were characterized by an immense data basis and hand computation rendered the obtaining of their solution infeasible. The electronic digital computer with its fast computational capability


for high volume of data became the most necessary tool in the applications of linear programming.

In June, 1949, a conference was held in Chicago by the Cowles Commission for Research in Economics. At this conference, a group of scientists of diverse backgrounds and specialization pooled their knowledge and their ideas to discuss the theory and practice of the most efficient utilization of limited world resources. There were three main sources of ideas, expressed in the conference, which led to the publication of some fundamental papers in the theory of linear programming: economists concerned with quantitative economics or econometrics (Wald, Schlesinger, von Neumann), welfare economics (Bergson, Pareto, Hotelling, and others) and specialists on inter-industry relationships or input-output techniques (Leontief, and others).

This conference laid the foundations of the dominating role that economics came to play in linear programming. Since then, many conferences have been held on linear programming or operations research and the amount of knowledge on the subject has increased geometrically.

In this chapter the concepts of optimization and mathematical programming in general and linear programming in particular will be clarified, some formulation and modeling rules will be presented, and a number of classical linear programming problems will be analyzed.

I. OPTIMIZATION AND MATHEMATICAL PROGRAMMING

For the first time, Euclid, in the third century B.C. presented a number of problems in his thirteen books of Elements, which could be called optimization problems.\(^\text{10}\) The theory of optimization studies the process of describing and attaining what is Best.\(^\text{11}\) The word \textit{optimum} means "best" either the quantitative "maximum" or "minimum" whatever the case may be. The verb \textit{optimize} means to achieve the optimum, and \textit{optimization} is the act of optimising. Therefore, \textit{optimization theory} "encompasses the quantitative study of optima and methods for finding them."\(^\text{11}\)

An example of Euclid's preoccupation with optimization was his effort to find the solution for the problem of drawing the longest and shortest straight line from a given point to the circumference of a given circle.

It was not until the eighteenth century with the development of the infinitesimal calculus and the calculus of variations that most of the optimization problems became amenable to rigorous solution methods.

The classical optimization models can be classified in


brief under the following categories:12

1. unconstrained optimization of a functional.
   Solution: by differential calculus;

2. constrained optimization of a functional.
   Solution: by use of Lagrangian multipliers;

3. unconstrained optimization of an integral.
   Solution: by calculus of variations;

4. constrained optimization of an integral.
   Solution: by use of Lagrangian multipliers and
   calculus of variations;

5. unconstrained optimization of several functionals.
   Solution: by the principle of Pareto optimality:13

6. constrained optimization of several functionals.
   Solution: by the principle of constrained Pareto
   optimality.

There are many subclasses to the above mentioned models
but in general all these classical models are characterized
by excellent mathematical analytical formulations and solu-
tions and absolutely weak numerical-computer application
capability.

12 Some good references on the subject would be:
Hancock, Harris, Theory of Maxima and Minima, (Dover Publications,
Inc., New York, 1960) and Forsyth, A.R., Calculus of Variations

13 Frisch, Ragnar, Maxima and Minima, (Rand McNally &
For a long time the planning of economic development and technological control could be carried out only by specialists in the field with suitable experience, education and intelligence. However, in recent years, with the development of large organizations with vast structural complexity, large-scale operation and distribution of resources and labor, the problem of planning and control became highly complicated. Decision making under these conditions necessitates the collection and processing of an enormous volume of information. Decisions sometimes involve the fate of a large sector of the economy and most of the time they have to be taken under fiscal constraints and resource limitations. Now, it is necessary to choose of all the possible ways, the most economical, which also gives the best results under the given constraints.

The classical optimization methods could conceptually solve problems of maximizing a functional of any number of variables under a set of constraints. Bertram Klein even extended the Lagragian-multiplier method to cover the non-negativity of the variables in problems with inequality constraints. But, the organizational problems of today are usually very large and such extensions of the classical


The methods proved to be very impractical except for the solution of "toy" problems. Therefore, suitable mathematical methods and techniques had to be devised for the solution of these problems. One of the most useful classes of the new optimization techniques developed to solve planning and control problems is called mathematical programming.

Mathematical programming refers to techniques of choosing the best schedule (program) of actions among alternatives in order to optimize a given functional or achieve a certain goal. If the objective function or goal and the constraints are linear then the method is called linear programming.

Linear programming problems are allocation problems in the sense that they combine activities and resources in such a way as to maximize total effectiveness.

Therefore, in general, the problems of mathematical programming could be given by the following model:

Optimize (Max. or Min.) \( Z = f(x_1, x_2, ..., x_n) \)
subject to
\[
\begin{align*}
    s_i(x_1, x_2, ..., x_n) &\leq 0 \quad (i = 1, 2, ..., m) \\
    x_j &\geq 0 \quad (j = 1, 2, ..., n)
\end{align*}
\]

where, the properties of the functions \( f \) and \( g \) characterize the type of mathematical programming model. For example, if \( f \) and \( g \) are linear then the model belongs to the class of linear programming.

Therefore, for mathematicians linear programming is the optimizing of a linear form subject to linear restraints. The economists define linear programming as a technique of allocating a group of limited resources among a number of competing demands. Other scientists have their own definitions which more or less qualify the abstract formulation of the mathematicians.

II. CURRENT STATE OF MATHEMATICAL PROGRAMMING

Mathematical programming problems can be divided into the following classes:\(^{17}\)

1. Continuous deterministic models: the decision vectors satisfy all constraints and define a connected feasible region; the functional to be optimized is continuous.

2. Discontinuous deterministic models: the feasible region is unconnected and (or) the objective functional is discontinuous.

3. Stochastic models: some or all of the coefficients in the constraints and in the objective functional undergo random variations.

4. Dynamic models: some or all of the coefficients in the constraints and in the objective functional are dependent on a parameter (e.g. the time).

5. Network models: the mathematical characteristics of these models are so special that major efficiencies in finding optimal solutions could be achieved by exploiting their special structural properties.

Linear programming can be found under the first class. Most of the fundamental research on this subject has essentially been completed, and only work on the computational aspects of the model is still continuing. The main emphasis of this work is directed towards the evolution of special methods for solving very large linear programming problems especially when they are characterized by special mathematical structure.

In this paper the mathematical foundations of linear programming will be given and certain computational aspects of the model will be analyzed. This approach is cardinal for the better development of linear programming which is the subject of this paper.


The stochastic models fall in general under the following categories:

1. Probabilistic programming.
2. Utility function approach.
3. Sequential approach.

Each one of these categories is subdivided into classes with different model characteristics and computational bias. Their analysis is reserved for later chapters.

III. FORMULATION OF A LINEAR PROGRAMMING MODEL

If an analysis is required for real-life systems either in existence or for purposes of design, the scientist is faced with a very complex problem. Usually the system under study is a complex aggregate of machines, people, supplies and facilities. There is an objective reason for the system's existence. For industry it may exist to produce different types of products or to distribute them in an economical manner, in the military it may exist to provide a defence force or a striking force for the carrier admirals.

The linear programming approach is to consider a system as decomposable into a number of components. Usually, there are three components about which information should be known.

before formulation occurs.

(a) The activities which are defined as those processes which combine commodities in definite proportions as inputs to produce outputs. The inter-activity relationships are represented by a series of equations in which the variables are the levels at which the activities are to be performed and the coefficients are these definite proportions which are the unit requirements of each activity for each commodity. These coefficients constitute the technology matrix of the system. Each equation is a continuity equation for every product in every place and every time period. Each equation presents the fact that the sum of the requirements of all the activities for a single resource cannot exceed the total amount available of that resource. If every continuity or material balance equation which describes the system is specified then a mathematical model of the process has been built.

(b) The resource availabilities are normally the availabilities of raw materials, market demands, capacities of plants, warehouses, manpower, transport fleets, etc.

(c) The objective function is the result of the qualification of the purpose of the model. There are many objectives of an organization which cannot be expressed in

Kent, Kenry, K., Linear Programming, (The National Cash Register Company, Dayton, Ohio, 1958) p. 10
mathematical forms. In such cases sometimes the function is replaced by a preference functional and the model is optimized for a preferred policy.

The diagram below illustrates the linear programming model.

System outputs: $x_i$ ($i=1,2,...,n$) levels of activity (which may be manufactured products or trained helicopter pilots)

System inputs: $b_j$ ($i=1,2,...,m$) capacity availability (which may be such tangible inputs as men, material, equipment or such intangibles as estimated capacities)

System technology: $A_{ij}$ ($i=1,2,...,n, j=1,2,...,m$) Matrix of technological coefficients or the transform matrix which turns the inputs into outputs (which may be manufacturing rates per different product or necessary time for training a paratrooper)

System constraints: $b_j$ ($j=1,2,...,m$) This is usually the same as system inputs plus the L.P. constraint of non-negativity for the $x_i$'s. ($x_i \geq 0, i=1,2,...,n$)

System objective: $z (c_i, x_j)$ to optimize.
Where: $z =$ measure of performance
$c_i =$ controllable variable: contribution $i$ to per unit of $j$
$x_j =$ uncontrollable variable: level of activity $j$

IV CHARACTERISTIC LIMITATIONS OF THE LINEAR PROGRAMMING MODEL

(1) Proportionality: The quantities of input and output flows of various items of the activity are always proportional to the activity level. For instance, if we wish to double the number of trained pilots coming out of a system, we would have to double the number of instructors and the number of draftees. a) $z$, and resource utilization must be proportional to the levels of activity $x$; b) Fixed charge problem: Let $D$ represent the incremental measure of resource utilization. Then we have

$$D = \begin{cases} 0 & \text{if } x \leq 0 \\ K + cx & \text{if } x > 0 \end{cases}$$

For $x > 0$ the set-up cost of a machine is $k$ but for $x = 0$ is zero. In this case $D$ is not linear over the whole range of $x$.  

values of \( x \) since \( x = 0 \), and therefore this problem does not conform to the Linear Programming model.

(2) \underline{Additivity}: For each kind of flow it is required that the total amount specified by the system as a whole equals the sum of amounts flowing into the various activities minus the sum of the amounts flowing out. Therefore for each kind of flow there is a characterization of a material balance equation. In other words, the additivity limitation does not permit nonlinearity to creep into the system because of interaction of activities.

(3) \underline{Divisibility}: The activity levels have to be divisible. If they are not the L.P. becomes an integer programming problem.

(4) \underline{Nonnegativity}: The activity levels have to be non-negative. Negative activity levels are not acceptable. The system can not train a negative number of pilots, or produce a negative amount of goods, which means that the product has to be stripped back to its raw materials!

(5) \underline{Deterministic constants}: \( a_{ij}, c_i, x_i, b; \) are all well known constants. If variability in the form of probability distributions or estimation errors is accepted, then the problem becomes a stochastic program.

(6) \underline{Linear objective function}: The problem amenable to solution by linear programming must have an exact quantitative objective. It is inadequate to state that "paper trim loss should be reduced." The objective of the problem has to
be given as a sum of terms each term being the product of a coefficient which might be the "cost" per unit of activity times this activity level.

Many problems which are non-linear can be made to fit into a linear programming model by a series of approximations. Special skill is required to break a non-linear function into line segments. If the effort is successful this technique is useful.

V. SOME LINEAR PROGRAMMING EXAMPLES

The following examples will give an idea of the characteristics of problems that can be attacked effectively with linear programming. The examples are realistic to the extent that industries do face such problems, but the problems presented are much smaller than the linear programming problems that are usually the center of industry models.

Since no numerical procedure for the solution of linear Programming problems has been presented yet, the solution will be given through some elementary geometrical concepts. Attention will be paid to the formulation of such problems.

(a) The Atlas Company manufactures two products, Product 1 and Product 2. Each Product 1 requires two hours in department A, four hours in department B and three hours in department C; each Product 2 requires four hours in department A and two hours in department B. The company's departments have variable processing capacity per week. Department
A can operate up to forty-eight hours, department B up to sixty hours and department C up to thirty-six hours. The company makes a profit of $6 on each Product 1 and $4 on each Product 2. There are no limitations in the number of Product 1 or Product 2 that the company can sell. The question that interests the management of the company is how to allocate the company's production capacity to maximize profit.

First the above problem statement has to be converted into a mathematical form by assigning the symbol $x_1$ to the number of Product 1 models manufactured in a week, and the symbol $x_2$ to the number of Product 2 models. The $x_1$ and $x_2$ are levels of activity in this production system.

The information about the problem can be formulated into a table. Table I states the problem in precise terms.

Table I. Atlas Manufacturing system

<table>
<thead>
<tr>
<th>Department</th>
<th>Time per Unit (Hrs)</th>
<th>Available Capacity (Hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Unit profit</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>
(a) **Formulation:** The problem can be analyzed as a system with the following Diagram 1.

---

**Diagram 1. ATLAS SYSTEM**

![Diagram](image)

Where we have as given:

- **System outputs:** level of activity $x_i$ ($i=1,2$)
- **System inputs:** capacity availability $b_n$ ($n=1,2,3$)
- **System technology:** time necessary for the manufacture of one unit of the products in each dept. (The central part of the given table for The Atlas Company)
- **System constraints:** available capacity and the general L.P. constraint of non-negativity for the $x_i$'s.
- **System objective:** maximum profit.
Now the problem can be formulated as a linear programming model:

Find values for $x_1, x_2$ in order to maximize $Z = 6x_1 + 4x_2$

Subject to:

\[
\begin{align*}
2x_1 + 4x_2 &\leq 48 \\
4x_1 + 2x_2 &\leq 60 \\
3x_1 &\leq 36 \\
\text{and} & \quad x_1 \geq 0 \\
& \quad x_2 \geq 0
\end{align*}
\]

There are two necessary steps for obtaining the graphical solution of this linear programming problem:

1. Plot the constraints.
2. Select the point in permissible values to maximize $z$.

The constraints are plotted in Figure 1. Then by giving different values to the objective function $z$, starting for example from $Z = 24 = 6x_1 + 4x_2$, several parallel lines are graphed by increasing $x_1, x_2$, and the maximum is found as $Z = 6(12) + 4(6) = 96$. In Figure 1 the shaded area constitutes the set of all feasible solutions, i.e., that do not violate any of the constraints. The vertices of the polygon ABCD constitute the set of all the basic feasible solutions (the concept will be clarified in the next chapter) and the vertex C is where the optimum solution is achieved.

(b) A car manufacturing company. Two cars (types) are manufactured and three different materials are being used. The material requirements for the manufacture of one
Figure 1. Graphical Solution of (a)
unit of each type of car are given by:

<table>
<thead>
<tr>
<th>Type of car</th>
<th>Available materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>rl</td>
<td>of steel</td>
</tr>
<tr>
<td>r2</td>
<td>of aluminum</td>
</tr>
<tr>
<td>4</td>
<td>1,000</td>
</tr>
<tr>
<td>5</td>
<td>of aluminum</td>
</tr>
<tr>
<td>2</td>
<td>1,000</td>
</tr>
<tr>
<td>3</td>
<td>of plastic</td>
</tr>
<tr>
<td>8</td>
<td>1,200</td>
</tr>
</tbody>
</table>

Profit of 5 units/unit car of type rl. Profit of 3 units/unit car of type r2.

\[ x_1 = \text{No. of units produced of type rl.} \]
\[ x_2 = \text{No. of units produced of type r2.} \]

The model can be formulated as:

Maximize Profit = 5x + 3x

Subject to:

\[ 4x_1 + 5x_2 \leq 1000 \]
\[ 5x_1 + 2x_2 \leq 1000 \]
\[ 3x + 8x_2 \leq 1200 \]
\[ x_1, x_2 \geq 0. \]

The graphical solution is given in Figure 2.

(c) A farmer buys feed for his cattle and he needs to supplement the vitamins in the feed. He has a choice of two different products, each of which contains the four vitamins required. These products contain the vitamins in different amounts. Like an intelligent farmer that he is,
Figure 2. Graphical Solution of problem (b)
he wants to meet the minimum vitamin requirements at the least possible cost. The data of his dilemma are given in Table

Table II. DATA OF FEED BLEND

<table>
<thead>
<tr>
<th>Vitamins (per pound)</th>
<th>Product 1 (units)</th>
<th>Product 2 (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin 1</td>
<td>250</td>
<td>50</td>
</tr>
<tr>
<td>Vitamin 2</td>
<td>170</td>
<td>250</td>
</tr>
<tr>
<td>Vitamin 3</td>
<td>200</td>
<td>150</td>
</tr>
<tr>
<td>Vitamin 4</td>
<td>50</td>
<td>500</td>
</tr>
<tr>
<td>Cost per pound</td>
<td>30 cents</td>
<td>40 cents</td>
</tr>
</tbody>
</table>

The farmer has to provide at least 1000 units of vitamin 1, 1500 units of vitamin 2, 1600 units of vitamin 3 and 2000 units of vitamin 4 for every hundred pounds of feed. Clearly, the farmer's objective function is \( z = 0.30x_1 + 0.40x_2 \) where \( x_1 \) is the number of pounds of product 1 purchased and \( x_2 \) represents the number of pounds for product 2. The constraints are the minimum requirements of the different vitamins that should be provided for every hundred pounds of feed. Each one of these four vitamins can be obtained either from product 1 or product 2 in varying amounts. But, whatever combination of products is bought, the total number of units for any given vitamin should equal or exceed the minimum requirement for that

\(^{23}\text{Adaptation from a problem given in: IBM, An Introduction to Linear Programming, (IBM, White Plains, N. Y., 1964) pp.7-8.}\)
vitamin. Therefore in mathematical form the constraints are:

\[
\begin{align*}
250x_1 + 50x_2 &\geq 1000 \\
100x_1 + 200x_2 &\geq 1500 \\
200x_1 + 150x_2 &\geq 1600 \\
50x_1 + 500x_2 &\geq 2000
\end{align*}
\]

Also, there are the additional constraints of non-negativity for the \(x_1\)'s and the \(x_2\)'s since negative buying is meaningless here.

The goal of the farmer is to meet the minimum requirements and at the same time not overspend. Therefore, the objective function is to minimize \(z = 0.30x_1 + 0.40x_2\) which is the total cost. In Figure 3 the constraints are plotted and the feasible region of operation is indicated. The feasible region in this problem is on the right-hand side of the coordinate system, since the constraints are "greater than or equal to" and there is no restriction on the amounts of product 1 or product 2 that the farmer can purchase. Also, there is no ceiling on the vitamin content of the feed-mix. The feasible region therefore consists of all combinations of products that do not violate any of the constraints. Now the objective function is plotted for different values of cost \(z\), and at the point \(a\), \(z\) achieves its maximum value. See Figure 3.
Figure 3. Feed Blend Constraint Plot
(d) A simple Input-Output Linear Programming Decision Model.\textsuperscript{23}

The AVA automobile manufacturing company wants to expand its production over a horizon of ten years. How the entire operation should be scheduled so as to maximize the total production of automobiles during this period? AVA Co. has two wholly owned subsidiaries: STE Steel Co. and TOO Tool Co. The function of STE is to apply AVA with steel for automobile production, and for factory building, and to supply TOO with steel for tool production and for possible expansion of tool factories. The function of TOO is to make tools for AVA, the steel mills of STE, and its own factories. Furthermore, AVA can buy fully equipped factories from Rose Corporation, which is on the verge of bankruptcy. Such a purchase would affect AVA's initial inventory position with regard to automobile factories, and it is desired to determine the effect of the purchase versus the number of the purchases on the expansion program. The problem is dynamic and the input to the \emph{i}th period is the output of the \emph{(i-1)st} period. The input to the first period is the initial inventory. The following are the variables on a per-period basis to be used in the problem:

\( x_1 = \) automobile production  
\( x_2 = \) steel production  
\( x_3 = \) tool production  
\( x_4 = \) automobile capacity production  
\( x_5 = \) steel capacity production  
\( x_6 = \) tool capacity production  
\( x_7 = \) storage of excess steel  
\( x_8 = \) storage of excess tools  
\( x_9 = \) storage of excess automobile capacity  
\( x_{10} = \) storage of excess steel capacity  
\( x_{11} = \) storage of excess tool capacity  
\( x_{12} = \) storage of excess people

**TABLE III. INPUT-OUTPUT TABLE**

<table>
<thead>
<tr>
<th>To make one unit of listed item per period</th>
<th>requires</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Steel</td>
<td>Tools</td>
<td>Auto cap.</td>
<td>Steel cap.</td>
<td>Tool cap.</td>
<td>people</td>
</tr>
<tr>
<td>autos</td>
<td>1.4</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>steel</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>tools</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>auto capacity</td>
<td>1</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>steel capacity</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>tool capacity</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>80</td>
</tr>
</tbody>
</table>

To distinguish between activity levels during each time period, a superscript \( i \) is added to each of the \( x \)'s. It is assumed that it takes one time period to create new capacity of any kind.
so that capacity created during period $i$ does not become available for production until the beginning of period $i+1$. It is clear that capacity will not be increased during the last period since the new capacity becomes available too late for the horizon planning. Therefore we have the following mathematical model:

$$\text{MAX } Z = \sum_{k=1}^{N} x_k^i$$

subject to:

steel constraint: $x_1^i + x_2^i = 1.4x_1^i + 1.2x_3^i + x_4^i + 3x_5^i + 2x_6^i + x_7^i$

tool constraint: $x_8^i + x_9^i = 3x_4^i + 6x_5^i + 2x_6^i + x_8^i$

auto capacity constraint: \[ \sum_{k=0}^{i-1} x_k^i = x_4^i + x_9^i \]

and \[ \sum_{k=0}^{i-1} x_k^i = x_2^i + x_10^i \]

labor constraint: \[ 16x_1^i + 21x_2^i + 15x_3^i + 17x_4^i + 180x_5^i + 80x_6^i + x_12^i = P \]

As each time period has six constraints, this problem deals with a system of $6N$ constraints. It is obvious that without the use of numerical-computer methods the problem will be very difficult to tackle geometrically. The solution of this problem will give the optimum scheduling of production of the auto, steel and tool companies for the production of the greatest number of automobiles. The actual solution will be given after the numerical simplex method is explained.
CHAPTER II

MATHEMATICAL THEORY OF LINEAR SPACES

INTRODUCTION

A number of mathematical disciplines contribute to the establishment of the foundations of linear programming. No attempt will be made here to give a complete coverage of these fields. Only those elements of these disciplines which have a direct affect on the linear programming theory will be discussed. No effort will be spared, however, for presenting a unified orderly development of the theory through a set-vector space-theoretic approach. The correct understanding of the bases of linear programming and its methods of solution hinges on the exact understanding of some of the mathematical constructs in n-dimensional vector space. That is where the emphasis is going to be placed. The stochastic linear programming theory is an extension of the deterministic approach given in the next chapter and it has the same foundations.

I. MATHEMATICAL CONSTRUCTS IN N-SPACE. GENERAL DEFINITIONS

Definition 2.1.1 A set $A$ is any given collection of objects which are called elements of $A$. 
In some cases a set can be determined by means of a property of its elements; for example, the set of positive rational numbers can be determined by the property that the elements \( x \) of the set are made out by the quotient of an integer \( p \) by an integer \( q \), where \( q \) is not zero.

If \( x \) is an element of the set \( A \), it is denoted by writing \( x \in A \) and conversely if \( x \) is not an element of \( A \) is written \( x \notin A \).

**Definition 2.2.2.** A set \( A_1 \) is called a **subset** of \( A \) if \( x \in A_1 \) implies that \( x \in A \). This is denoted as \( A_1 \subseteq A \), meaning that set \( A \) is contained in set \( A \). Two sets \( A_1 \) and \( A_2 \) are **equal** if \( A_1 \subseteq A_2 \) and \( A_2 \subseteq A_1 \).

**Definition 2.2.3.** A **null set** is a set having no elements and is denoted by \( \emptyset \). The null set is a subset to every set.

The set consisting of the elements \( a_1, a_2, a_3, \ldots, a_n \) is denoted by \( [a_1, a_2, a_3, \ldots, a_n] \). The set of elements of \( A \) which satisfy a property \( P(a) \) is denoted by

\[
S = \{a \in A \mid P(a)\}
\]

The above notation reads "\( S \) is the set of all elements in \( A \) for which the property \( P \) holds."

**Definition 2.2.4.** The **union** of the two sets \( S_1 \) and \( S_2 \), written \( S_1 \cup S_2 \), is the set \( S_3 = \{ x \mid x \in S_1 \text{ or } x \in S_2 \} \)

The **intersection** of the two sets \( S_1 \) and \( S_2 \), written as \( S_1 \cap S_2 \), is the set \( S_3 = \{ x \mid x \in S_1 \text{ and } x \in S_2 \} \).

**Definition 2.2.5.** The two sets \( S_1 \) and \( S_2 \) are said to be **disjoint** if \( S_1 \cap S_2 = \emptyset \).
Usually the set of positive integers $[0, 1, 2, \ldots, n, \ldots]$ is denoted by $\mathbb{N}$ and the set of real numbers is denoted by $\mathbb{R}$. $\mathbb{R}$ can be represented by means of the points on an oriented line with an origin $0$, in such a way that to the number $a$, there corresponds the point $P$ whose abscissa is $OP = a$.

**Definition 2.2.6.** The difference $S_1 - S_2$ set for given $S_1$ and $S_2$ is the set $S_3 = \{ x \in S_1 \mid x \notin S_2 \}$.

**Definition 2.2.7.** $\mathbb{R}$ is called the real or Euclidean line.

**Definition 2.2.8.** The Cartesian product $S_1 \times S_2$ of two sets $S_1$ and $S_2$ is defined to be the set formed by the pairs $(x_1, x_2)$, where $x_1 \in S_1$, and $x_2 \in S_2$. Therefore, $\mathbb{R}_1 \times \mathbb{R}_2$ can be represented by the points of a plane determined by two oriented rectangular axes. This plane as presented in Figure 2.1 is called the Euclidean plane, and is denoted by $\mathbb{R}^2$.

![Figure 4.E - plane](image-url)
Definition 2.1.9. The product \( \mathbb{R}^n = \mathbb{R} \times \mathbb{R} \cdots \mathbb{R} = \{ (x_1, x_2, \ldots, x_n) | x_1, x_2, \ldots, x_n \in \mathbb{R} \} \) which is the set of \( n \)-tuples of real numbers, is called Euclidean space of \( n \) dimensions.

Definition 2.1.10. An ordered set is a set in which the notions of "larger", "smaller", "maximum", "minimum" etc. are defined.

Definition 2.1.11. A nonempty set of elements \( G \) is called a group if in \( G \) there is defined a binary operation, called the product and denoted by \( \cdot \) such that:

1. \( x_1, x_2 \in G \) implies that \( x_1 \cdot x_2 \in G \)
2. \( x_1, x_2, x_3 \in G \) implies that \( x_1 \cdot (x_2 \cdot x_3) = (x_1 \cdot x_2) \cdot x_3 \)
3. There exists an element \( e \in G \) such that \( x_1 \cdot e = e \cdot x_1 = x_1 \) for all \( x_1 \in G \)
4. For every \( x \in G \) there exists an element \( x^{-1} \in G \) such that \( x \cdot x^{-1} = x^{-1} \cdot x = e \).

Definition 2.1.12. A group \( G \) is said to be abelian if for every \( x_1, x_2 \in G \), \( x_1 \cdot x_2 = x_2 \cdot x_1 \).

Definition 2.1.13. An \( n \)-vector is an ordered set of \( n \)-tuples of real numbers called components such that

\[
\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R}
\]

Definition 2.1.14. A vector space \( V \) over a field \( F \) is a set of elements called vectors, such that any two vectors
$\bar{x}_1$ and $\bar{x}_2$ of $V$ determine a unique vector $\bar{x}_1 + \bar{x}_2$ as sum, and any $\bar{x} \in V$ and $a \in F$ determine a product $a \cdot \bar{x} \in V$ such that:

1. $V$ is an abelian group under addition
2. $a \cdot (\bar{x}_1 + \bar{x}_2) = a \cdot \bar{x}_1 + a \cdot \bar{x}_2$, $(a_1 + a_2)$
   \[ \cdot \bar{x} = a_1 \cdot \bar{x} + a_2 \cdot \bar{x} \]
3. $(a_1, a_2) \cdot \bar{x} = a_1 \cdot (\bar{x} \cdot a_2)$, $1 \cdot \bar{x} = \bar{x}$

The set of all $n$-vectors, structured by the above operations constitute the vector space $\mathbb{R}^n$ defined in Definition 2.2.9. By fixing a reference system of oriented rectangular axes with origin $0$, every $n$-vector $[x_1, x_2, \ldots, x_n]$ corresponds to the point with coordinates $x_1, x_2, \ldots, x_n$ and conversely; the origin $0$ corresponds to the null vector. Therefore, associated with the vector space $\mathbb{R}^n$ there is a space of points which is called the affine space. There is exact interchangeability between the affine and the vector space over the field $\mathbb{R}$ of the real numbers; and this allows easy translation of reasoning from the algebraic to geometric and conversely which will prove invaluable in the continuing development of the foundations of linear programming. The mixed affine-vector concepts allow extended flexibility in the exposition.

**Definition 2.2.15.** A subspace of a vector space $V$ is a subset of $V$ which is itself a vector space with respect to the operations of addition and scalar multiplication in $V$.

**Definition 2.2.16.** A null vector $\bar{0}$ of $\mathbb{R}^n$ is a vector for which all components are zero. i.e. $\bar{0} = [0, 0, \ldots, 0]$. A unit vector $\bar{e}_1$ of $\mathbb{R}^n$ is the vector whose $i$th component
is +1, all others being zero, i.e. \( \bar{e}_2 = 0,1,0,\ldots,0 \).

**Definition 2.2.17.** Two vector subspaces, \( R_1 \) and \( R_2 \), are said to be supplementary if

\[
R_1 + R_2 = \mathbb{R}, R_1 \cap R_2 = 0
\]

**Definition 2.2.18.** A subset \( S_1 \) of \( \mathbb{R}^n \) is called a linear variety if

\[
\begin{align*}
\tilde{x}_1, \tilde{x}_2 &\in S_1 \\
a_1, a_2 &\in \mathbb{R} \\
a_1 + a_2 = 1
\end{align*} \implies a_1 \tilde{x}_1 + a_2 \tilde{x}_2 \in S_1
\]

(2.3)

![Diagram](image)

**Figure 5. Linear variety in \( S_1 \subset \mathbb{R}^3 \)**

**Theorem 2.2.1.** If \( E \) is a vector subspace and \( \bar{x}_0 \) is a fixed vector in \( \mathbb{R}^n \), the set \( L = E + \bar{x}_0 \) is a linear variety. Conversely, every linear variety is of the form \( L = E + \bar{x}_0 \) for some vector subspace \( E \) and some fixed vector \( \bar{x}_0 \).

**Proof:** \( E + \bar{x}_0 \) is a linear variety, for if \( a_1 \) and \( a_2 \) are real numbers such that \( a_1 + a_2 = 1 \), we have
$$\overline{x}_1, \overline{x}_2 \in E + \overline{x}_0 \implies \begin{cases} \overline{x}_1 - \overline{x}_0 \in E \\ \overline{x}_2 - \overline{x}_0 \in E. \end{cases} \implies$$

(2.4)

$$\implies (a_1 \overline{x}_1 + a_2 \overline{x}_2) - \overline{x}_0 = a_1 (\overline{x}_1 - \overline{x}_0) + a_2 (\overline{x}_2 - \overline{x}_0) \in E$$

$$\implies a_1 \overline{x}_1 + a_2 \overline{x}_2 \in E + \overline{x}_0$$

Now suppose that $A$ is a linear variety; it shall be proved that there exists a vector subspace $E$ and a fixed vector or a point $\overline{x}_0$ such that $A = E + \overline{x}_0$.

If $A = \emptyset$, then $A = \emptyset = \emptyset + \overline{x}_0$.

If $A \neq \emptyset$, let $\overline{x}_0$ be a point in $A$ and write $E = A - \overline{x}_0$, then $E$ is a vector subspace, for

$$\begin{align*}
(1) \quad \overline{x}_1 \in E & \implies \begin{cases} \overline{x}_1 = \overline{y}_1 - \overline{x}_0 \\ \overline{y}_1 \in A \end{cases} \implies a_1 \overline{x}_1 = a_1 \overline{y}_1 - a_1 \overline{x}_0 = a_1 \overline{y}_1 + (1-a_1) \overline{x}_0 \in A - \overline{x}_0 = E, \quad (2.5) \\
(2) \quad \overline{x}_1, \overline{x}_2 \in E & \implies \begin{cases} \overline{x}_1 = \overline{y}_1 - \overline{x}_0 \\ \overline{y}_1 \in A \\ \overline{x}_2 = \overline{y}_2 - \overline{x}_0 \end{cases} \implies \overline{x}_1 + \overline{x}_2 = \\
&= \frac{\overline{y}_1 + \overline{y}_2}{2} - \overline{x}_0 \in A - \overline{x}_0 \\
&= \overline{x}_1 + \overline{x}_2 \in E. \quad (2.6) \\
\therefore \quad A = E + \overline{x}_0, \text{ where } E \text{ is a vector subspace.}
\end{align*}$$

If $A$ is a linear variety such that $A = E_1 + \overline{x}_1$ and $A = E_2 + \overline{x}_2$, then $E_1 = E_2$. Because $\overline{x}_1 - \overline{x}_2 \in E_2$

since $\overline{x}_1 = 0 + \overline{x}_1 \in A = E_2 + \overline{x}_2$

then $E_1 = E_2 + (\overline{x}_2 - \overline{x}_1) \subseteq E_2 + E_2 \subseteq E_2$ \quad (2.7)

By symmetry, $E_2 \subseteq E_1$ and therefore $E_1 = E_2$. 


Definition 2.1.19. The space $E$ such that $A = E + x_0$ (necessarily unique according to theorem 2.1.1) is called the subspace parallel to $A$. If two linear varieties $A$ and $B$ have the same parallel subspace $E$, they are called parallel to one another.

A linear variety $S$ which had a straight line $Sy$ through the point $0$ as a parallel subspace is called a straight line. If a linear variety $P$ has a plane through the point $0$ as a parallel subspace it is called a plane.

II. VECTOR OPERATIONS. DEPENDENCE. RANK. DIMENSION. BASES

Definition 2.2.1. The scalar multiple of an $n$-vector by a scalar number $a_1$ is an $n$-vector formed by multiplying each component by $a_1$. Therefore, for a column vector

$$
\begin{pmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_n
\end{pmatrix}
\quad
\begin{pmatrix}
    a_1 x_1 \\
    a_1 x_2 \\
    \vdots \\
    a_1 x_n
\end{pmatrix}
$$

(2.8)

Definition 2.2.2. To every pair of vectors $\overline{x}$ and $\overline{y}$ in $\mathbb{R}^n$ there corresponds a real number called the inner product of $\overline{x}$ and $\overline{y}$ written

$$
\overline{x} \cdot \overline{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n
$$

(2.9)

and such that

(a) $\overline{x} \cdot \overline{y} = \overline{y} \cdot \overline{x}$

(2.10)

(b) $(a_1 \overline{x} + a_2 \overline{y}) \cdot \overline{z} = a_1(\overline{x} \cdot \overline{z}) + a_2(\overline{y} \cdot \overline{z})$ (2.11)

for all $a_1, a_2 \in \mathbb{R}$ and $\overline{x}, \overline{y}, \overline{z} \in \mathbb{R}^n$. 


The vectors \( \mathbf{x} \) and \( \mathbf{y} \) are called orthogonal if their inner product is equal to zero.

**Definition 2.2.1.** The sum of two \( n \)-vectors is the vector formed by adding the corresponding components. Thus

\[
\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}
\]

**Definition 2.2.4.** Two \( n \)-vectors are equal if their corresponding components are equal. Thus

\[
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{implies} \quad x_i = y_i \quad \text{for} \quad i = 1, 2, \ldots, n.
\]

**Definition 2.2.5.** The vectors \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \) are said to be linearly independent if and only if

\[
\sum_{j=1}^{k} a_j \mathbf{x}_j = \mathbf{0} \quad \Rightarrow \quad a_j = 0 \quad (j = 1, \ldots, k)
\]

otherwise, they are called dependent.

**Definition 2.2.6.** A vector \( \mathbf{x} \) is a linear combination of the vectors \( \mathbf{x}_j \) \((j = 1, \ldots, k) \) if

\[
\mathbf{x} = \sum_{j=1}^{k} a_j \mathbf{x}_j
\]

The null vector is a linear combination of all vectors, if \( a_j \) is taken equal to zero. This is a trivial case and it will not be considered in the study of vector dependency.
For further clarification of the vector space concepts an example is given.

Suppose that a description of the U. S. production of certain industrial and agricultural goods, such as automobiles, wheat, and so forth is wanted. The required information for each state can be given by an ordered sequence of \( n \) real numbers, the \( i \)th number being the quantity of the \( i \)th product which the state produces. Therefore, given Table IV,

**TABLE IV**

**ANNUAL PRODUCTION IN TONS**

<table>
<thead>
<tr>
<th></th>
<th>Wheat</th>
<th>Steel</th>
<th>Coal</th>
<th>...</th>
<th>Electronics</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>( a_{11} )</td>
<td>( a_{12} )</td>
<td>( a_{13} )</td>
<td>...</td>
<td>( a_{1n} )</td>
</tr>
<tr>
<td>State 2</td>
<td>( a_{21} )</td>
<td>( a_{22} )</td>
<td>( a_{23} )</td>
<td>...</td>
<td>( a_{2n} )</td>
</tr>
<tr>
<td>State 3</td>
<td>( a_{31} )</td>
<td>( a_{32} )</td>
<td>( a_{33} )</td>
<td>...</td>
<td>( a_{3n} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>State ( M )</td>
<td>( a_{m1} )</td>
<td>( a_{m2} )</td>
<td>( a_{m3} )</td>
<td>...</td>
<td>( a_{mn} )</td>
</tr>
</tbody>
</table>

it can be said that each year state 2 produces \( a_{21} \) tons of wheat, \( a_{22} \) tons of steel, ..., \( a_{2n} \) tons of electronics.

Now, if the total productivity for the year is required for two states, simply add the vectors of the individual states' productivity:

\[
\text{State 1 } \overrightarrow{p_1} = (a_{11}, a_{12}, a_{13}, \ldots, a_{1n}) \\
\text{State 2 } \overrightarrow{p_2} = (a_{21}, a_{22}, a_{23}, \ldots, a_{2n})
\]

therefore, combined production is \( \overrightarrow{p_1} + \overrightarrow{p_2} = (a_{11} + a_{21}, a_{12} + a_{22}, a_{13} + a_{23}, \ldots, a_{1n} + a_{2n}) \).
If the productivity in each commodity increases by a factor $k \in \mathbb{R}$ in state 1, then the new productivity will be:

$$\text{New } \overline{p}_1 = k\overline{p}_1 = (ka_{11}, ka_{12}, ka_{13}, \ldots, ka_{1n}).$$

**Definition 2.2.7.** The rank of a subset $S$ of a vector space $V$ is the maximal number of independent vectors in $S$. If $S$ is itself a vector space, $S \subseteq V$, the rank of $S$ is called the number of dimensions of the space $S$, or simply the dimension of $S$. The dimension of a linear variety is, by definition, equal to the dimension of the subspace parallel to this variety.

**Definition 2.2.8.** A basis of an $r$-dimensional vector space $V^r$ (or of a subset $R^r$ or rank $r$, of a space $V$) is called any set of $r$ linearly independent vectors of $V^r$ (or $R^r$) such that all other vectors in the space can be generated as linear combinations of the vectors in the set.

**Theorem 2.2.1.** A sufficient condition for $(r + 1)$ vectors $\overline{y}_i (i = 1, \ldots, r+1)$ of $V$ to be linearly dependent is that each of these vectors be a linear combination of the same set of $r$ vectors $\overline{x}_j (j = 1, \ldots, r)$ of $V$.

**Proof:** The proof will use mathematical induction and in the trivial case where either one or several of the $\overline{y}_i$ is the vector $\overline{0}$ will be discounted.

Let

$$\overline{y}_i = \sum_{j=1}^{r} a_{ij} \overline{x}_j, \quad i = 1, \ldots, (r+1)$$

(2.17)

The theorem is true for $r=1$ because $a_{11}$ and $a_{21}$, being different from zero (since neither $\overline{y}_1$ nor $\overline{y}_2$ is 0) give

$$a_{11}\overline{y}_2 - a_{21}\overline{y}_1 = 0$$

(2.18)
Suppose that the theorem is true for \( r = k - 1 \); it will be shown that it is true for \( r = k \). Therefore,

\[
\overline{y}_i = \sum_{j=1}^{k} a_{ij} \overline{x}_j \quad i = 1, \ldots, (k + 1) \tag{2.19}
\]

and suppose \( a_{11} \neq 0 \) and set

\[
\overline{z}_i = \overline{y}_i - \frac{a_{i1}}{a_{11}} \overline{y}_1 = \sum_{j=2}^{k} \left( a_{ij} - \frac{a_{i1}}{a_{11}} \right) \overline{x}_j \quad i = 2, \ldots, (k + 1).
\tag{2.20}
\]

Now, since \( \overline{z}_i \) depends linearly on the same \( (k - 1) \) vectors \( \overline{x}_j \) \((j = 2, \ldots, k)\) the hypothesis permits

\[
\overline{0} = \sum_{i=2}^{k+1} a_i \overline{z}_i = \sum_{i=2}^{k+1} a_i \overline{y}_i - \left( \sum_{i=2}^{k+1} a_{i1} \right) \overline{y}_1 \tag{2.21}
\]

with at least one of the \( a_i \neq 0 \), which proves the theorem.

**Corollary 2.2.1.1.** \( \mathbb{R}^n \) is an \( n \)-dimensional space.

**Proof:** The \( n \) unit vectors \( \overline{e}_i \) \((i = 1, \ldots, n)\) are linearly independent since

\[
\sum_{i=1}^{n} a_i \overline{e}_i \quad (a_1, a_2, \ldots, a_n) = \overline{0} \Rightarrow a_1 = a_2 = \ldots = a_n = 0
\]

and every vector \( \overline{x} = (x_i) \) may be written

\[
\overline{x} = \sum_{i=1}^{n} x_i \overline{e}_i \tag{2.22}
\]

which proves by theorem 2.2.1 that \( (n + 1) \) arbitrary vectors of \( \mathbb{R}^n \) are linearly dependent.

The conclusion from this corollary and definition 2.2.8 is that every set of \( n \) linearly independent vectors of \( \mathbb{R}^n \) constitutes a basis of the \( \mathbb{R}^n \) space.
Theorem 2.2.2. A subset of \( r \) non-zero linearly independent vector \( x_i \) of a set \( S \) of vectors is a basis of \( S \) if and only if every vector of \( S \) is a linear combination of the \( x_i(i=1,\ldots,r) \).

**Proof:** If every vector \( x \) of \( S \) is a linear combination of the \( x_i(i=1,\ldots,r) \) according to theorem 2.2.1(r+1) arbitrary vectors of \( S \) are dependent, therefore \( S \) is of rank \( r \) and \( x_i \) constitutes a basis, by definition.

Conversely, if the set \( x_i(i=1,\ldots,r) \) constitutes a basis, \( S \) is of rank \( r \), by definition, and for every \( y \in S \), the vectors \( y \) and \( x_i(i=1,\ldots,r) \) are dependent.

Theorem 2.2.3. Let \( y \) be any vector in the vector space generated by a set of independent vectors \( x_i(i=1,\ldots,r) \); then the representation of the \( y \) as a linear function of the set \( x_i(i=1,\ldots,r) \) and the values \( a_i(i=1,\ldots,r) \) such that

\[
a_1x_1+a_2x_2+\ldots+a_rx_r=y
\]

(2.23)

**Proof:** If the representation is not unique then there exists another set of values \( b_i(i=1,\ldots,r) \) such that

\[
b_1x_1+b_2x_2+\ldots+b_rx_r=y
\]

(2.24)

If substraction is applied (2.24) from (2.23) the result is

\[
(a_1-b_1)x_1+(a_2-b_2)x_2+\ldots+(a_r-b_r)x_r=0
\]

(2.25)

and the conclusion is that if not all \( a_i-b_i=0 \), the vectors \( x_1,x_2,\ldots,x_r \) are not independent. This conclusion is contrary to the assumption that \( x_i(i=1,\ldots,r) \) are independent, and therefore the theorem is proven.
The expression (2.23) is the representation of a vector $\overline{y}$ in terms of the basis $\overline{x}_i$ ($i = 1, \ldots, r$) and the $a_i$ ($i = 1, \ldots, r$) are the coordinates of $\overline{p}$ relative to this basis.

**Theorem 2.2.4.** Given a set of vectors $\overline{x}_i$ ($i = 1, \ldots, r$) which being linearly independent constitute a basis, and a vector $\overline{p} \neq 0$ in a vector space, it is possible to replace one of the columns of the basis by $\overline{p}$ to form a new basis.

**Proof:** Let the representation of $\overline{p}$ in terms of the basis be

$$c_1\overline{x}_1 + c_2\overline{x}_2 + \cdots + c_r\overline{x}_r = \overline{p} \quad (2.26)$$

Now, since $\overline{p} \neq 0$ at least one $c_i \neq 0$ in (2.26). If $c_1 \neq 0$ then it will be shown that a new basis can be formed by replacing $\overline{x}_1$ by $\overline{p}$. According to theorem 2.2.3, the representation (2.26) is unique and therefore $\overline{x}_1$, $\overline{x}_2$, $\ldots$, $\overline{x}_r$, $\overline{p}$ are linearly independent because if not the implication is that $\overline{p}$ has a non-zero coefficient and thus can be represented differently in terms of the other vectors from the unique representation (2.26) and that constitutes a contradiction to theorem 2.2.3.

Therefore, it only need to be shown that an arbitrary $\overline{y}$ can be expressed in terms of $\overline{x}_2$, $\overline{x}_3$, $\ldots$, $\overline{x}_r$, $\overline{p}$ to prove that these vectors form a basis. Therefore, multiplying (2.26) by a constant $k$ and subtracting from (2.23) the result is

$$(a_1 - kc_1) \overline{x}_1 + (a_2 - kc_2) \overline{x}_2 + \cdots + (a_r - kc_r) \overline{x}_r + \overline{p}k = \overline{y} \quad (2.27)$$

Therefore if $k = \frac{a_1}{c_1}$ ($c_1 \neq 0$) \((2.28)\)

then the conclusion is that $\overline{y}$ is linearly dependent on the
vectors $\overline{x}_2, \ldots, \overline{x}_r, \overline{p}$, sine $a_1 - kc_1 = 0$. Therefore, the new basis $\overline{x}_2, \ldots, \overline{x}_r, \overline{p}$ can generate any other vector $\overline{y}$ in the space.

**Theorem 2.2.5.** Given a set of vectors $\overline{x}_i$ ($i=1, \ldots, r$) which being linearly independent constitute a basis and 1 independent non-zero vectors $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_l$ in the vector space generated by the basis, if is possible to replace 1 vectors in the basis by $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_l$.

**Proof:** The proof is given by mathematical induction when $l = 1$ according to theorem 2.2.4 the theorem 2.2.5 stands. Now, suppose that it is true that a new basis can be formed by substituting 1-1 vectors $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_l$ for 1-1 vectors in the basis by replacing $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_l$. Therefore, the new basis is $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_l, \overline{x}_1, \ldots, \overline{x}_r$. Now let the representation of $\overline{q}_1$ in terms of this basis be

$$c_1 \overline{q}_1 + c_2 \overline{q}_2 + \ldots + c_{l-1} \overline{q}_{l-1} + c_l \overline{x}_1 + \ldots + c_r \overline{x}_r = \overline{q}_1. \quad (2.29)$$

At least one $c_i \neq 0$, for $i \neq 1$, because otherwise $\overline{q}_1$ would be linearly dependent on $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_{l-1}$ which is contrary to the assumption of independency of the $\overline{q}_i$ ($i=1, \ldots, l$). Therefore, let $c_k \neq 0$ for some $k \neq 1$. Then according to the argument of theorem 2.2.4, $\overline{q}_1$ can replace $\overline{q}_k$ in this basis to form a new basis consisting of vectors $\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_{l-1}, \overline{q}_{l}, \overline{x}_{l+1}, \ldots, \overline{x}_r$.

### III ELEMENTS OF THE THEORY OF MATRICES AND DETERMINANTS

Problems concerning groups of vectors in $\mathbb{R}^n$ and their operations center around the fundamental concept of linear
transformations in $\mathbb{R}^n$ and the concept of matrices.

**Definition 2.3.1.** A rectangular array of scalar quantities, called elements, having $m$ rows and $n$ columns, is called an $m \times n$ matrix and is arranged in the following manner:

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix} \quad (2.30)
\]

**Definition 2.3.2.** The following are true for the matrix $A$ in $(2.30)$:

(a) The $i$-th row of $A$ is the ordered $n$-tuple $(a_{i1}, a_{i2}, \ldots, a_{in})$ and according to definition 2.2.13 it is a vector, called row-vector;

(b) The $j$-th column of $A$ is the ordered $m$-tuple $(a_{1j}, a_{2j}, \ldots, a_{mj})$ and according to definition 2.2.13 it is a vector, called column-vector;

(c) If $m = n$ matrix $A$ is called a square matrix;

(d) If $a_{ij} = 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ then matrix $A$ is called a zero matrix;

(e) If $a_{ij} = 1$ for all $i = j$ and $a_{ij} = 0$ for all $i \neq j$ then matrix $A$ is called an identity matrix and is denoted by $I$.
(f) If \( a_{ij} = k_i, k \in \mathbb{R} \) for all \( i = j \) and \( a_{ij} = 0 \) for all \( i \neq j \) then matrix \( A \) is called a \textit{diagonal} matrix;

(g) The transpose of \( A \) is a matrix whose \((i,j)\) element is the \((j,i)\) element of \( A \). The transpose is denoted by \( A' \);

(h) A square matrix is called \textit{symmetric} if it is equal to its transpose i.e. \( A = A' \); \textit{skewsymmetric} if it is the negative of its transpose i.e. \( A = -A' \);

(i) A square matrix is called \textit{triangular} if all \( a_{ij} = 0 \) for \( i < j \) or if all \( a_{ij} = 0 \) for \( i > j \);

(j) Associated with every square matrix \( A \) is a number called the \textit{determinant} of \( A \) denoted by \(|A|\). The determinant \( A \) is the polynomial

\[
|A| = \sum (-1)^d a_{d_1} d_2 \ldots d_n,
\]

where \((d_1, d_2, \ldots, d_n)\) is a permutation of the \( n \) first integers and \( d \) is the number of inversions of this permutation. The summation is taken over all possible permutations of the second indices. Therefore, the determinant is a homogeneous polynomial in the \( a_{ij} \) and it has \( n! \) terms. The determinant has the following properties:

1. \(|A| = 0\), if every \( a_{ij} = 0 \) for \( i = 1, \ldots, n; j = 1, \ldots, n \);

2. The value of the determinant is not changed if corresponding columns and rows are interchanged;

3. If \(|D|\) is the determinant formed by
interchanging two columns or rows in $|A|$, then $|D| = -|A|$;

(4) If two columns or rows of $|A|$ are identical, then $|A| = 0$;

(5) If every element of a row or of a column of a determinant is multiplied by a constant $k$, then the value of the determinant is multiplied by $k$;

(6) The value of a determinant is not changed if, to every element of a column or row it is added $k$ times the corresponding element of another column or row;

(7) $|AB| = |A| \cdot |B|$, $A$ and $B$ are $n \times n$ matrices;

(8) For every $p$ and $q$, between 1 and $n$

$$\sum_{j=1}^{n} a_{pj} |A_{qj}| = |A| S_{pq},$$

where: $S_{pq} = 1$ if $p = q$

$S_{pq} = 0$ if $p \neq q$

If $p = q$ the above relation is called the expansion of $|A|$ by its $p$th row;

(9) If the columns or rows of a determinant are linearly dependent, this determinant is zero. This is a deduction from properties (1) and (6);

(10) In 2-space, corresponding to two column-vectors $\overline{a}_1 = (a_{11}, a_{21})$ and $\overline{a}_2 = (a_{12}, a_{22})$ there is the determinant
whose absolute value is equal to the area of the parallelogram with sides \( \vec{a}_1 \) and \( \vec{a}_2 \) (Fig. 6).

In 3-space, the three linearly independent vectors

\[
\vec{a}_1 = (a_{11}, a_{21}, a_{31}), \quad \vec{a}_2 = (a_{12}, a_{22}, a_{32}), \quad \vec{a}_3 = (a_{13}, a_{23}, a_{33})
\]

determine a parallelepiped (Fig. 7) and the absolute value of the determinant
48

\[ \begin{vmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{vmatrix} \]

is equal to the volume of the parallelepiped;

(k) A square matrix A is called **nonsingular** if \( |A| \neq 0 \). If \( |A| = 0 \), the matrix is singular;

(l) The **minor** \( D_{ij} \) of the element \( a_{ij} \) of a matrix A is the determinant obtained from the square matrix A by striking out the \( i \)th row and \( j \)th column;

(m) The **cofactor** \( A_{ij} \) of the element \( a_{ij} \) equals \((-1)^{i+j} D_{ij}\). \( A_{ij} \) is called the **signed minor** of the element \( a_{ij} \);

(n) The **adjoint matrix** of A is a matrix \( A^+ \) whose \((i,j)\) element is the \((j,i)\) element of the matrix whose element \((i,j)\) is the cofactor of \( a_{ij} \) in \( |A| \).

**Definition 2.3.3.** The following operations are valid in matrices. Given matrices \( A, B, C \) and \( a, a_1, a_2 \in R \)

(a) Two matrices A and B are **equal** when
\[ a_{ij} = b_{ij} \quad i = 1, \ldots, m; \quad j = 1, \ldots, n; \]

(b) The sum of two matrices A and B is defined as
\[ A + B = (a_{ij} + b_{ij}) = C = C_{ij} \quad (2.31) \]
and it has the following properties
\[ A + B = B + A \]
\[ (A + B) + C = A + (B + C) \]  \hspace{1cm} (2.32)

(c) The product of a matrix by a real number is defined by
\[ aA = Aa = (aa_{ij}) \]
and with the sum it has the following properties
\[ A + (-1)A = (0) \]
\[ (a_1 + a_2)A = a_1 A + a_2 A \]  \hspace{1cm} (2.33)
\[ a(A + B) = aA + aB \]

(d) The product of a matrix \( A, mxp \), with a matrix \( B, pxn \), is the matrix \( C, mxn \), defined by the relation
\[ c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj} \hspace{1cm} i = 1, \ldots, m; \]
\[ j = 1, \ldots, n \]  \hspace{1cm} (2.34)

Therefore, \( AI_n = A = I_mA \)  \hspace{1cm} (2.35)
The product of two matrices in the order \( AB \) is defined only if the number of columns of \( A \) is equal to the number of rows of \( B \). If \( A \) and \( B \) are square matrices of the same order, the two products \( AB \) and \( BA \) are defined, but lead to two different results. The matrix product is not commutative, but is associative and distributive with respect to summation, and also possesses the property of homogeneity with respect to
scalar multiplication, \( A(aB) = a(AB) \)

(e) The transpose has the following properties

\[
(A^T)^T = A \\
(A + B)^T = A^T + B^T \\
(AB)^T = B^T A^T
\]

Every \((mxp)\) matrix may be considered as a column of \(m\) row-vectors, or as a row of \(n\) column-vectors. Therefore, two matrices \(A\) and \(B\) can be written

\[
A = \begin{bmatrix}
- \\
a_1 \\
\vdots \\
\vdots \\
\bar{a}_m
\end{bmatrix}: \bar{a}_i \text{ being a row vector}
\]

\[
B = \begin{bmatrix}
b_1 \\
\vdots \\
\vdots \\
b_n
\end{bmatrix}: \bar{b}_j \text{ being a column vector}
\]

The product \(AB\) can be written

\[
AB = A \begin{bmatrix}
b_1 \\
\vdots \\
\vdots \\
b_n
\end{bmatrix} = \begin{bmatrix}
\bar{a}b_1 \\
\bar{a}b_2 \\
\vdots \\
\bar{a}b_n
\end{bmatrix} = (2.37)
\]

or given that \(A\) and \(B\) are \((nxn)\) square matrices

\[
BA = \begin{bmatrix}
b_1 \\
\vdots \\
\vdots \\
b_n
\end{bmatrix} \begin{bmatrix}
\bar{a}_1 \\
\bar{a}_2 \\
\vdots \\
\bar{a}_n
\end{bmatrix} = \bar{b}_1 \bar{a}_1 + \cdots + \bar{b}_n \bar{a}_n = (2.38)
\]
Definition 2.3.4. If all rows \( k \) and all columns but \( s \) are crossed out of an \((m \times n)\) matrix \( A \), the result \((k \times s)\) matrix is called a submatrix of \( A \).

Definition 2.3.5. A partition of a matrix \( A \) is the division of \( A \) into submatrices such that

\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix} = \begin{bmatrix}
c_{11} & c_{12} \\
c_{21} & c_{22}
\end{bmatrix}
\tag{2.39}
\]

Most of the operations applicable to the elements of \( A \) are applicable to the submatrices \( C_{11}, C_{12}, C_{21}, C_{22} \) taken as elements of \( A \).

Definition 2.3.5(a). The rank of a matrix is the maximal number of column-vectors or row-vectors of the matrix which are linearly independent.

Definition 2.3.6. A square matrix of order \( n \) is called regular if its rank is \( n \); otherwise it is called singular.

Theorem 2.3.1. The rank of the columns of a matrix is the same as the rank of its rows.

Proof: Suppose that the first \( r \) columns and the first \( s \) rows are linearly independent. These linearly independent vectors from a \((s \times r)\) submatrix designated \( A_1 \).
It has to be shown that \( r = s \). Since the vector \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_s \) constitutes a basis in the space \( \mathbb{R}^s \) according to theorem 2.2.4 one of the vectors of the bases can be replaced by a non-zero vector in the vector space to form a new basis. Assume that vector \( \bar{y}_r = \sum k_i \bar{x}_i \) and then \( k_i \neq 0 \). Therefore, by theorem 2.2.4 \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{s-1}, \bar{y}_r \) is a basis for \( \mathbb{R}^s \).

Now, it can be written
\[
\bar{y}_{r-1} = \sum_{i=1}^{s-1} e_i \bar{x}_i + \bar{y}_r
\]
and at least one \( e_i \) must be non-zero, because otherwise the set of \( \bar{y}_i \) would not be linearly independent. Take \( e_{s-1} \neq 0 \); then \( \bar{x}_1, \ldots, \bar{x}_{s-2}, \bar{y}_{r-1}, \bar{y}_r \) is a basis for \( \mathbb{R}^s \). Now this process can continue until a basis is found which must take one of the following forms:
\[
\bar{x}_1, \ldots, \bar{x}_{s-r}, \bar{y}_1, \ldots, \bar{y}_r \text{ or } \bar{y}_1, \ldots, \bar{y}_r
\]
There should be at least as many \( \bar{x}_i \) as \( \bar{y}_j \), because otherwise a state would be reached where one \( \bar{y}_j \) was a linear combination of some of the others which contradicts the independence assumption. Therefore \( s \geq r \). Now, if one can start with the \( \bar{y}_j \) and insert the \( x_i \) the result would be \( r \geq s \). Therefore \( s = r \) and the theorem is proven.

**Theorem 2.3.2.** The rank of a matrix \( A \) is \( r \) if and only if there exists at least one square submatrix of \( A \) of
order r whose determinant is non-zero, and if all square submatrices of A of order larger than r have a zero determinant.

**Proof:** If matrix A is of rank r then (r + k) arbitrary column-vectors of A are linearly dependent and therefore every square submatrix of order (r + k) has a zero determinant according to property (9) of Definition 2.3.2 section (j).

Let \( \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_r \) be r linearly independent columns of A. According to the proof of theorem 2.3.1, these vectors form an m\( \times \)r matrix \( A(\bar{r}) \) with \( m \geq r \). Now suppose that all the square submatrices of order r taken from \( A(\bar{r}) \) have a zero determinant. Consider especially the determinants of the \((m - r + 1)\) matrices of order r which have as \((r - 1)\) first rows the \((r - 1)\) first rows of \( A(\bar{r}) \), and let \( |A_{rj}| \) be the cofactor of element \( a_{rj} \) in the determinant which is formed from the r first rows of \( A(\bar{r}) \). Now, by hypothesis

\[
\sum_{j=1}^{r} a_{rj} |A_{rj}| = 0 \quad (4.42)
\]

\[
\sum_{j=1}^{r} a_{ij} |A_{ij}| = 0 \quad i = (r + 1), \ldots, m \quad (4.43)
\]

Moreover,

\[
\sum_{j=1}^{r} a_{ij} |A_{rj}| = 0 \quad i = 1, \ldots, (r-1) \quad (4.44)
\]

since formula (2.44) is just the expansion of a determinant which has two identical rows of indices i and r. Now, it can be deduced from (2.42), (2.43) and (2.44) that there exists the same relation of linear dependence between the elements
of each of the rows of \( A(r) \). But, this contradicts the hypothesis and proves that the condition is necessary.

For the condition to be sufficient, suppose that every determinant of order larger than \( r \) is zero and that there exists at least one non-zero determinant of order \( r \). A cannot be of rank \( p < r \) because from definition 2.3.5 about the rank of a matrix and property (9) of definition 2.3.2 section (j), every determinant of order \( r \) would be zero. And from the necessary condition part of the proof, A cannot be of order \( l > r \) because there would exist at least one non-zero determinant of order \( l \). Therefore, the condition is sufficient.

**Definition 2.3.6.** Given a nonsingular \( nxn \) matrix \( A \), it is called the **inverse matrix** \( A^{-1} \) of \( A \) the matrix, which, if it exists satisfies the relations

\[
A^{-1}A = AA^{-1} = In
\]  

\[(2.45)\]

**Theorem 2.3.3.** The inverse \( A^{-1} \) of a nonsingular matrix \( A \) is unique.

**Proof:** According to definition 2.3.2 section (n) about matrix \( A^+ \) and section (j) no. (8) about the expansion of \( |A| \)

\[
A^+A = AA^+ = |A| In
\]  

\[(2.46)\]

and, consequently

\[
\frac{A^+}{|A|} = A^{-1}
\]  

\[(2.47)\]

Now, in order to prove uniqueness, suppose that there exists another matrix \( B \) or \( C \), such that
BA = I_n
AC = I_n

Then B = BAA^{-1} = I_nA^{-1} = A^{-1}
C = A^{-1}AC = A^{-1} I_n = A^{-1}

**Definition 2.3.7.** Product form of the inverse. Let $A$ be an mxn nonsingular matrix for which the inverse $A^{-1}$ is known. Let $A$ be written in terms of its column-vectors as

$$A = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p, \ldots, \bar{a}_m)$$  \hspace{1cm} (2.48)

Now, since matrix $A$ is nonsingular, the vectors $\bar{a}_j$ ($j = 1, \ldots, m$) are linearly independent and therefore form a basis. A new matrix $A_1$ can be deduced from $A$ by substitution of a vector $\bar{p}$ for the vector $\bar{a}_p$. This gives a new basis according to theorem 2.2.4. The vector $\bar{p}$ can be expressed in terms of the vectors $\bar{a}_j$ according to theorem 2.2.2. Then, if $\bar{y}$ is the vector expressing $\bar{p}$ as a function of the vectors $\bar{a}_j$ the following expression stands

$$\bar{p} = Ay = \sum_{j=1}^{m} y_j \bar{a}_j$$  \hspace{1cm} (2.49)

The proposition is to find the inverse of

$$A_b = (\bar{a}_1, \bar{a}_2, \ldots, \bar{p}, \ldots, \bar{a}_m)$$  \hspace{1cm} (2.50)

A necessary and sufficient condition such that $A_1^{-1}$ exists is that $|A_1| \neq 0$ or that the vectors $(\bar{a}_1, \bar{a}_2, \ldots, \bar{p}, \ldots, \bar{a}_m)$ are linearly independent or that $y_p \neq 0$. Suppose that this is true. Then the following expression can be written.

$$\bar{a}_p = \frac{1}{y_p} \sum_{j \neq p} y_j a_j + \frac{1}{y_p} \bar{p}$$  \hspace{1cm} (2.51)

so that $\bar{a}_p = A_b \bar{p}$  \hspace{1cm} (2.52)
Where:

\[
\bar{Z}_p' = \left[ -\frac{y_1}{y_r}, \frac{y_2}{y_r}, \ldots, -\frac{y_{p-1}}{y_r}, 1, -\frac{y_{p+1}}{y_r}, \ldots, -\frac{y_m}{y_r} \right]
\]  \hspace{1cm} (2.53)

Now, since the product of an arbitrary matrix with the
1th unit vector is the 1th column of the matrix the combination
of (2.48), (2.50) and (2.52) will give

\[ A = A_b \left( E_1, E_2, \ldots, \bar{E}_{p-1}, \bar{Z}_p, \bar{Z}_{p}, \bar{Z}_{p+1}, \ldots, E_m \right) = A_b J \]  \hspace{1cm} (2.54)

The matrix \( J \) is obtained from the unit matrix \( I_m \) where
the \( p \)th column \( E_p \) has been replaced by \( \bar{Z}_p \).

If (2.54) is multiplied on the left by \( A_b^{-1} \) and on the
right by \( A^{-1} \) the result is

\[ A_b^{-1} = J_p A^{-1} \]

The preceding result can be used now for a programmed
calculation of \( A^{-1} \) of a nonsingular \( mxm \) matrix

\[ A = (\tilde{a}_1, \ldots, \tilde{a}_m) \]

Step 1. Introduce unit matrix

\[ I_m = (\bar{E}_1, \bar{E}_2, \ldots, \bar{E}_m) \]

Step 2. Replace \( \bar{E}_1 \) by \( \tilde{a}_1, \bar{E}_2 \) by \( \tilde{a}_2 \), etc.

\[ A_1 = (\tilde{a}_1, \bar{E}_2, \ldots, \bar{E}_m) \]
\[ A_2 = (\tilde{a}_1, \tilde{a}_2, \ldots, \bar{E}_m) \]
\[ A_m = (\bar{E}_1, \bar{E}_2, \ldots, \tilde{a}_m) \]

and

\[ A_1, A_2, \ldots, A_m = A \]
Step 3. \[ A_1^{-1} = J_1 I_m = J_1 \]
\[ A_2^{-1} = J_1 A_1^{-1} = J_2 J_1 \]
\[ A^{-1} = A_m^{-1} = J_m A^{-1} = J_m J_{m-1} \cdots J_1 \]  (2.55)

**Definition 2.4.7.** The inverse matrix A has the following properties subject to easy verification.

- (1) \((A^{-1})^{-1} = A\)
- (2) \((A^t)^{-1} = (A^{-1})^t\)
- (3) \((AB)^{-1} = B^{-1}A^{-1}\)
- (4) \(AB = 0 \quad A \neq 0 \quad B = 0\)
- (5) \(A^{-1} = \frac{1}{A}\)

**IV. LINEAR TRANSFORMATIONS**

**Definition 2.4.1.** If \(S_1\) and \(S_2\) are nonempty sets, then a mapping from \(S_1\) to \(S_2\) is a subset, \(M\), of \(S_1 \times S_2\) such that for every \(s_1 \in S_1\) there is a unique \(s_2 \in S_2\) such that the ordered pair \((s_1, s_2)\) is in \(M\). \(s_2\) is called the image of \(s_1\) under the mapping.

**Definition 2.4.2.** A linear transformation \(T\) is a mapping of a vector space \(R^n\) into a vector space \(R^m\), which has the following two properties:

- \(T(\overline{x}_1, \overline{x}_2) = T(\overline{x}_1) + T(\overline{x}_2)\)  (2.56)
- \(T(a\overline{x}) = aT(\overline{x}), \quad a \in R\)  (2.57)

The linear transformation, denoted by \(T\), is an operator which applied to an arbitrary vector of \(R^n\), transforms it into a vector of \(R^m\).

**Theorem 2.4.1.** A real \(m \times n\) matrix \(A\) completely defines a linear transformation of \(R^n\) into \(R^m\). The linear transform-
mation of $\mathbb{R}^n$ into $\mathbb{R}^m$. The linear transformation is equivalent to the product of $A$ with the vectors of $\mathbb{R}^n$.

Proof: Let $(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n)$ and $(\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_m)$ be the unit bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Any arbitrary vector $\vec{x} \in \mathbb{R}^n$ and its transformation $\vec{y} = T(\vec{x})$ can be exposed as functions of the respective bases, such that

$$\vec{x} = \sum_{j=1}^{n} x_j \vec{e}_j \quad (2.58)$$

$$\vec{y} = \sum_{i=1}^{m} y_i \vec{f}_i \quad (2.59)$$

The transformation $T(\vec{e}_j) \in \mathbb{R}^m$ can be expressed as a function of $\vec{f}_i$.

$$T(\vec{e}_j) = \sum_{i=1}^{m} a_{ij} \vec{f}_i \quad j = 1, \ldots, n \quad (2.60)$$

Now, if (2.58) and (2.60) are combined the result is

$$\vec{y} = T(\vec{x}) = \sum_{j=1}^{n} x_j T(\vec{e}_j)$$

$$\vec{y} = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} x_j \right) \vec{f}_i \quad (2.61)$$

and by comparison with (2.59) and from theorem 2.3.3

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \quad (2.62)$$

so that $\vec{y} = Ax$, where $A = (a_{ij})$, $i = 1, \ldots, m$, $j = 1, \ldots, n$.

The matrix $A$ is defined uniquely and independently of $\vec{x}$ from (2.60) and therefore it defines the given linear transformation. From (2.32), (2.33), and (2.34) it can be seen also that the product of an $m \times n$ matrix $A$ with a vector of $\mathbb{R}^n$ or with a $1 \times n$ matrix of $\mathbb{R}^n$ is a linear transformation of $\mathbb{R}^n$ into $\mathbb{R}^m$.

Definition 2.4.3. The rank of a linear transformation is the dimension of the subspace $S$ of $\mathbb{R}^m$ transformed from
Theorem 2.4.2. The rank of a linear transformation $T$ is equal to the rank of the matrix $A$ which defines $T$.

**Proof:** In (2.62) the column-vectors $\bar{a}_j$ of $A$ are the vectors transformed from the unit vectors $\vec{e}_j$ $(j = 1, 111, n)$. Then it follows from definition 2.2.7 about the rank of a subset and the dimension of a vector space, that the dimension of $S$ of $\mathbb{R}^m$ is equal to the maximal number of linearly independent column-vectors $\bar{a}_j$, in other words, is equal to the rank of $A$.

**Definition 2.4.4.** The transformation $T_3$ defined by 

$$T_3 (x) = T_2 (T_1(x))$$

is called the product of the two linear transformations $T_1$ and $T_2$.

**Definition 2.4.5.** Elementary transformations are certain manipulations which when carried out on the rows or columns of a matrix have the property of replacing it by a matrix of the same rank.

The elementary transformations are:

(i) Addition of one row to another row

(ii) Multiplication of a row by a non-zero scalar

(iii) The exchange of two rows

The importance of the elementary transformations derives from the fact that the Gaussian elimination procedure for the calculation of the solutions of a system of linear equations depends on these transformations.
CHAPTER III

CONVEX SPACES AND LINEAR PROGRAMMING

INTRODUCTION

This chapter deals with properties of convex sets of points in an n-dimensional point space \( \mathbb{R}^n \). The fundamental theorems of linear programming and duality are explained and presented in mathematical form. These theorems constitute the bases from where stochastic linear programming extends itself.

I. BASIC PROPERTIES OF CONVEX SETS

Definition 3.1.1 If \( \bar{a} \in \mathbb{R}^n \) and \( \bar{a} \neq 0 \), the line passing through \( 0 \) and generated by \( \bar{a} \) is called the set of points,

\[ \Delta = \{ x \mid x = \lambda \bar{a}, \lambda \in \mathbb{R} \} \]

a linear variety parallel to a line passing through \( 0 \) is called a line.

Definition 3.1.2 The ray generated by \( \bar{a} \) is the set of points.

\[ D = \{ x \mid x = \lambda \bar{a}, \lambda > 0 \} \]

Definition 3.1.3 If \( \bar{a} \in \mathbb{R}^n \), \( \bar{a} \neq 0 \) and \( \alpha \in \mathbb{R} \) then the set

\[ P = \{ x \mid \bar{a}^t x = \alpha \} \]

is called a hyperplane (Fig. 8) and the set

\[ H = \{ x \mid \bar{a}^t x > \alpha \} \]

is called a closed half-space. (Fig. 8)

A hyperplane divides \( \mathbb{R}^n \) into two closed half spaces.
and they are usually denoted by
\[
H^+ = (\bar{x} | \bar{a} \cdot \bar{x} > \alpha)
\]
\[
H^- = (\bar{x} | \bar{a} \cdot \bar{x} < \alpha)
\]

\(\bar{a}\) is interpreted as a row-vector and \(\bar{x}\) as a column-vector.
For \(\bar{x} = (1, 2, 3)\), \(\bar{a} = (\bar{a}_1, \bar{a}_2, \bar{a}_3)\) and \(\alpha = 1\). The inner product is
\[
\bar{a} \cdot \bar{x} = \bar{a}_1 + 2\bar{a}_2 + 3\bar{a}_3
\]
and the two half-spaces are
\[
\bar{a}_1 + 2\bar{a}_2 + 3\bar{a}_3 > 1
\]
\[
\bar{a}_1 + 2\bar{a}_2 + 3\bar{a}_3 < 1
\]
as shown in Figure 8. The points on the plane \(\bar{a}_1 + 2\bar{a}_2 + 3\bar{a}_3 = 1\) belong to both half-spaces.

![Figure 8. Illustration of a hyperplane in \(R^3\)](image)

When a set \(S\) is situated on one side of a hyperplane \(\bar{a} \cdot \bar{x} = b\), one of the half spaces \(\bar{a} \cdot \bar{x} \geq b\) or \(\bar{a} \cdot \bar{x} \leq b\) contains the \(S\); this half-space is then called a support of, or a supporting half-space for \(S\).

A hyperplane is called a supporting hyperplane \(P_s\) to a convex set \(C\) if it contains at least one point of \(C\) and such that the points of \(C\) are on one side of \(P_s\).
Definition 3.1.4. A cone $G$ is a set of points or vectors such that 
$$\bar{x} \in G \Rightarrow \lambda \bar{x} \in G$$

In other words, a cone is a set which, if it contains a vector $\bar{x}$, contains also the ray generated by $\bar{x}$.

Definition 3.1.5. A set $C$ is said to be convex if

$$\text{for } \lambda_1, \lambda_2 \in \mathbb{R}, \quad \begin{cases} \bar{x}_1, \bar{x}_2 \in C \\ \lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 \in C \end{cases} \Rightarrow \lambda_1 + \lambda_2 = 1$$

According to this definition, every linear variety is convex. In $\mathbb{R}^n$, for a given $f$ numerical function not identically zero, the following sets are convex:

- The hyperplane $P = \{ \bar{x} \mid f(\bar{x}) = \alpha \}$
- The open half-space $H^- = \{ \bar{x} \mid f(\bar{x}) < \alpha \}$
- The closed half-space $H^- = \{ \bar{x} \mid f(\bar{x}) \leq \alpha \}$

Geometrically, a set $C$ is convex if for each pair $(\bar{x}_1, \bar{x}_2)$ of points of $C$, the linear segment having these two points for extremities is contained in the set.

Figure 9 shows some convex and non-convex sets.

Theorem 3.1.1. A linear convex combination of $K$ points $\bar{x}_i$, is a point defined by 
$$\bar{x} = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \ldots + a_p \bar{x}_p$$

where $a_i \in \mathbb{R}, a_i \geq 0$, and $\sum a_i = 1$.

A set $C$ is convex if every linear convex combination of the points of $C$ is also a point of $C$. Therefore for any $P$
$$x_i \in C, (i=1, \ldots, p)$$
Figure 9. Convex sets in $\mathbb{R}^3$.
\[ a_\zeta > 0 \prod_{\zeta=1}^{\zeta=p} \sum_{z=1}^{\zeta=p} a_\zeta x_\zeta \in \mathcal{C} \quad (3.4) \]

**Proof:** The proof is given by induction. According to definition 3.1.5 the theorem is true for \( p=2 \). Now, assume that the theorem is true for \( p=k \) and prove that it is true for \( p=k+1 \). If the theorem is true for \( k \) then set

\[ \bar{x} = \sum_{\zeta=1}^{k+1} a_\zeta \bar{x}_\zeta \quad a_\zeta > 0 \quad \sum_{\zeta=1}^{k+1} a_\zeta = 1 \]

If \( a_{k+1} = 1 \) then \( \bar{x} = \bar{x}_k \in \mathcal{C} \)

If \( 1 > a_{k+1} \) then \( \bar{x} = (1-a_{k+1}) \sum_{\zeta=1}^{k} \frac{a_\zeta}{1-a_{k+1}} \bar{x}_\zeta + a_{k+1} \bar{x}_{k+1} \)

\( \bar{x}_c \) is a point of \( \mathcal{C} \) by hypothesis and therefore \( \bar{x} \) is also a point of \( \mathcal{C} \).

**Definition 3.1.6** The **convex closure** \( \mathcal{S} \) of a set \( S \) is the intersection of all convex sets containing \( S \). Usually, the term **convex hull** is applied for the convex closure of a finite set \( (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) \). Convexity is preserved under set intersection:

**Theorem 3.1.2** If \( A \) and \( B \) are convex subsets of \( \mathbb{R}^n \), then intersection \( A \cap B \) is also a convex subset of \( \mathbb{R}^n \).

**Proof:** Suppose that \( \bar{x}_1 \) and \( \bar{x}_2 \) are points of \( A \cap B \), then \( \bar{x}_1, \bar{x}_2 \in A \) and \( \bar{x}_1, \bar{x}_2 \in B \) and the linear segment connecting \( \bar{x}_1 \) and \( \bar{x}_2 \) belongs to both \( A \) and \( B \) because both \( A \) and \( B \) are assumed to be convex. This completes the proof of the theorem.

**Definition 3.1.7** Any point, at a finite distance, which cannot be represented as a linear convex combination with
non-zero coefficients of two distinct points of a convex set $C$, is called an extremal point. In other words every point-vector $\bar{x}$ such that it is impossible to find two points $\bar{x}_1$ and $\bar{x}_2$ in $C$ satisfying,

$$\bar{x} = a \, \bar{x}_1 + (1-a) \, \bar{x}_2, \quad 0 < a < 1, \quad a \in \mathbb{R}$$

The set of all the extremal points of a convex set $C$ is called the profile of $C$. When a non-empty convex set $C$ is closed and bounded, classical results of calculus state that the linear function $\bar{a}x$ possesses a maximum and a minimum on $C$, in other words it attains its least upper bound $M$ and its greatest lower bound $m$ for at least one value $\bar{x}_1$ or $\bar{x}_2$ of $\bar{x}$, so that

$$\begin{align*}
\bar{a} \bar{x}_1 &= M = \sup (\bar{a}x) \quad (3.5) \\
\bar{a} \bar{x}_2 &= m = \inf (\bar{a}x) \quad (3.6)
\end{align*}$$

The equations $\bar{a} \bar{x} = M$ and $\bar{a} \bar{x} = m$ represent from (3.5) and (3.6), two supporting hyperplanes to the set $C$.

**Theorem 3.1.3** A convex set $C$ in $\mathbb{R}^n$, closed and bounded from below (or from above), possesses at least one extremal point; every supporting hyperplane to $C$ contains at least one extremal point of $C$.

**Proof:** If $C$ is a convex set, $P_s$ a supporting hyperplane to $C$, and $T = C \cap P_s$, by Theorem 3.1.2, is a convex set then every extremal point $\bar{x}$ of $T$ is also an extremal point of $C$, because if $\bar{x}$ is not an extremal point of $C$ then

$$\bar{x} = a \, \bar{x}_1 + (1-a) \, \bar{x}_2, \quad \bar{x}_1, \bar{x}_2, \in C, \quad 0 < a < 1 \quad (3.7)$$
By definition the equation of $P_s$ is $\bar{a}x=b$ and
\[\bar{a}x_1=b\]
\[\bar{a}x_2=b\]
and since $\bar{x}$ is in $P_s$ and from (3.7)
\[b=\bar{a}x=\bar{a}x_1+(1-a)\bar{a}x_2\]
and therefore $\bar{a}x_1=\bar{a}x_2=b$, which contradicts the hypothesis that $\bar{x}$ is an extremal point of $T$. Now, if every closed and bounded from below convex set $C^{n-1}$ of $R^{n-1}$ possesses an extremal point, and if $C^n$ is a closed and bounded from below convex set is $R^n$, every supporting hyperplane $P$ to $C^n$ contains an extremal point of $C^n$; it suffices to apply the preliminary preceding result to $C^n \cap P_s$. $C^n$ possesses at least one extremal point since there exists at least one supporting hyperplane to $C^n$; the theorem is true for $N=2$ since a closed and bounded from below convex set of $R$ is a half-line and obviously possesses at least one extremal point.

**Definition 3.1.8** A cone is a convex cone if it is a convex set. A convex set spanned by a finite number of points is a convex polyhedron, and a convex polyhedral cone is a cone generated by a convex polyhedron.

**Theorem 3.1.4** A convex polyhedral cone is the collection of all linear combinations
\[
\left( \sum_{j=1}^{n} \lambda_j \bar{p}_j \right) \lambda_j > 0
\]
where $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n$ are fixed points.
Clearly this is fulfilled for the constraints in the formal statement of any linear programming problem.

**Definition 3.1.2** A simplex is an n-dimensional convex polyhedron having exactly N+1 vertices. The boundary of the simplex contains simplices of dimension lower than n which are called simplicial faces. A simplex of zero dimension is a point; of one dimension it is a line; of two dimensions it is a triangle and of three dimensions it is a tetrahedron.

**Theorem 3.1.5** If C is a convex set and \( x_k \) is a point not in the closure of 0; \( (C) \), there is a hyperplane separating \( x_k \) from \( (C) \).

**Theorem 3.1.6** If there is a non-empty set \( S \) of vectors \( (x) \), the set \( T \) of vectors \( y \) such that \( (x,y) \) for every \( x \in S \) is a closed convex set.

**Definition 3.1.10** Given that \( C \) is a closed convex cone, the set of points \( y \) such that \( (x,y) \geq 0 \) for every \( x \in C \) is called the dual space \( C^* \), and \( C^* \) is a closed convex cone.

**Theorem 3.1.7** If \( C \) is a closed convex cone and \( C^* \) is its dual, then \((C^*)^* = C\).

24 Most of the proofs that are not given here, they can be found in: Marshall, Hall, Jr., Combinatorial Theory (Blaisdell, Waltham, Massachusetts, 1967) pp.66-90.
Theorem 3.1.8 (Farkas) Let $A$ be an $m \times n$ matrix and suppose that $\bar{y} = (y_1, \ldots, y_n)$ is a vector such that $\bar{y}w' = (\bar{y}, \bar{w})' \cdot c$ for every vector $\bar{w} = (w_1, \ldots, w_n)$ such that $Aw' \geq 0$. Then $\bar{y}$ is of the form $\bar{y} = xA$, where $x \geq 0$.

II THE GENERAL LINEAR PROGRAMMING PROBLEM

Linear programming is concerned with the problem of minimizing or maximizing a linear function subject to constraints on linear functions of the vector variable $\bar{x}$. In any given problem the constraints can be given by linear equations, linear inequalities, or a mixture of the two. Since $A\bar{x} = \bar{b}$ is equivalent to the two systems $A\bar{x} \geq \bar{b}$ and $A\bar{x} \leq \bar{b}$, it is always possible to express all constraints in a given problem as inequality constraints. The general linear programming problem can be stated in the following form:

**Linear Program P:**

Max $Z = \bar{c}' \cdot \bar{x}$  \hspace{1cm} (3.10)

subject to $A\bar{x} \leq \bar{b}$  \hspace{1cm} (3.11)

and $\bar{x} \geq 0$  \hspace{1cm} (3.12)

This linear program has a corresponding dual linear program which can be stated as:

**Linear Program D:**

Min $Z' = \bar{b}' \bar{y}$  \hspace{1cm} (3.13)

subject to $A'\bar{y} \geq \bar{c}$  \hspace{1cm} (3.14)

and $\bar{y} \geq 0$  \hspace{1cm} (3.15)
The dual program is constructed from the primal by changing the sense of the inequalities, transposing the matrix $A$ and interchanging the positions of the vectors $\bar{C}$ and $\bar{b}$.

Some definitions will be given now and the most important theorems characterizing the solution to the general linear programming problems will be stated, without proofs, which can be found in every textbook on linear programming. 25

**Definition 3.2.1** A feasible solution to a linear program is a vector $\bar{X}$ which satisfies conditions (3.11) and (3.12).

**Definition 3.2.2** A basic feasible solution is a feasible solution with no more than $m$ positive $X_i$ if the matrix $A$ of (3.11) is a nxm matrix.

**Definition 3.2.3** An optimum feasible solution is a feasible solution which also maximizes (3.10) or minimized (3.13).

**Theorem 3.2.1** The set of all feasible solutions to the linear-programming problem is a convex set, $K$.

**Theorem 3.2.2** The objective function (3.10) assumes its maximum at an extremal point of the convex set $K$. (3.13) assumes its minimum at an extremal point of $K$ also. If the objective function assumes its minimum at more

than one extremal points, then it takes on the same value for every convex combination of those particular points.

Theorem 3.2.3  If a set of $K \leq M$ vectors $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_k$ can be found that are linearly independent and such that

$$x_1 \bar{a}_1 + x_2 \bar{a}_2 + \ldots + x_k \bar{a}_k = \bar{a}_0$$

and all $x_i \geq 0$, then the point $\bar{x} = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)$ is an extremal point of the convex set of feasible solutions.

Theorem 3.2.4  If $\bar{x}$ is an extreme point of $K$, then the vectors associated with positive $X_i$ form a linearly independent set. From this it follows that, at most, $M$ of the $X_i$ are positive.

Theorem 3.2.5  Associated with every extreme point of $K$ is a set of $M$ linearly independent vectors from the given set $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n$.

Theorem 3.2.6  $\bar{x}$ is an extremal point of $K$ if and only if the positive $X_j$ are coefficients of linearly independent vectors $\bar{a}_j$ in

$$\sum_{j=1}^{n} X_j \bar{a}_j = \bar{a}_0$$

Theorem 3.2.7  The linear programs $P$ and $D$ possess solutions if and only if both problems are feasible. Furthermore, if $Z_0$ is the maximum value of $C'\bar{x}$ in program $P$ and $Z_0'$ is the minimum value of $B'\bar{y}$ in program $D$, then $Z_0 = Z_0'$. If either problem has a solution, so does the other (theorem of duality).
III THE SIMPLEX PROCEDURE

The simplex procedure is the basic computational procedure for finding the optima in linear programming problems. Other procedures that have been developed are extensions of the simplex method. The simplex procedure starts from any basic (extremal point) feasible solution and obtains an optimum feasible solution in a number of finite steps. These steps or iterations consist in finding a new feasible solution whose corresponding value of the objective function is more that the value of the objective function for the preceding solution. The iterations continue until a maximum solution has been reached. According to Theorem 3.2.4, all extremal solutions have \( M \) linearly independent vectors associated with them. Therefore, the search for the optimum linear program should be confined to those programs that are generated by \( M \) linearly independent vectors. In other words, only the vertices of the convex set \( K \) of the constraints should be examined.

The ideas behind the simplex method can be understood through a representation of the action of the method in \( 3 \)-dimensional space which can be easily constructed.

Consider the linear programming problem:

\[
\begin{align*}
\text{Max} \quad & z = c_1 x_1 + c_2 x_2 + c_3 x_3 \\
\text{subject to} \quad & (E_1) \quad a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \leq b_1
\end{align*}
\]
The geometrical interpretation of the preceding inequalities can be given within the three-dimensional space in which \( x_1, x_2 \) and \( x_3 \) denote the coordinates of a point relative to orthogonal axes. Constraint \((E_1)\) defines a half-space bounded by the plane

\[
a_{11} x_1 + a_{12} x_2 + a_{13} x_3 = b_1
\]

The region consisting of all points that satisfy each of the constraints \((E_1), (E_2), \ldots, (E_p)\) is an intersection of half-spaces. If this region is empty, the linear program is inconsistent and has no solution. By inconsistent is meant that there is not one single point that satisfies its constraints. A non-empty intersection of half-spaces can be a convex polyhedron, polygon, line segment, or a point and can also be bounded or unbounded. For geometrical simplification we limit the results of the non-empty intersection of half-spaces to the Figure 10.
The problem now then is to find the highest extremal point of such a convex polyhedron. The highest point is not necessarily unique since a whole edge of the polyhedron might have the same height.

The polyhedron consists of faces, edges and vertices. Face 1 in figure 10. is part of the plane

\[ a_{11} x_1 + a_{12} x_2 + a_{13} x_3 = b_1 \]

Face 2 is part of the plane

\[ a_{21} x_1 + a_{22} x_2 + a_{23} x_3 = b_2 \]

and so forth. The line segment PQ is an edge and it is the intersection of three faces, namely faces 1, 2, and 3.

The simplex method gives an efficient algorithm for examining the vertices to find the optimum solution.

The first step in the simplex method is to find a vertex on the polyhedron. This can be done easily if a set of \( K \) vectors can be found that are linearly independent since according to the Theorem 3.2.3, this will give an extremal point on the convex polyhedron. Now, suppose that the extremal point \( P \) has been found. Then according to Figure 10 the point \( P \) is the intersection of planes 1, 2, and 3. The first three inequalities \((E_1), (E_2)\) and \((E_3)\), therefore, determine this point.

The second step is to find the intersection of planes 1 and 2, then move from \( P \) along this edge and determine
whether the value of the objective function is increasing or decreasing. The same procedure is carried out for planes 2 and 3 and also for planes 3 and 1. If no increase of the value of the objective function occurs, then P is a maximizing point for the linear program.

The third step is to follow along an edge which increases the value of the objective function, which in Figure 10 is the edge defined by planes 2 and 3, and terminate at the extremal point S.

At the extremal point S the three steps are repeated and this leads to another vertex which increases the value of the objective function, in this case it is vertex U. Then from U the procedure leads to vertex W which gives the maximum value to the objective function, and thus solves the problem. These geometrical instructions are translated into arithmetical action procedures in the example that follows.
A simple machine loading model for clarification and qualification of the simplex method in the primal and dual of linear programming.

<table>
<thead>
<tr>
<th>Dual</th>
<th>Primal</th>
<th>Maximize</th>
<th>Capacities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>$W_1$</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$W_2$</td>
<td></td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Criterion ($$)</td>
<td>$W$</td>
<td>$W$</td>
<td>Solution</td>
</tr>
<tr>
<td></td>
<td>$C_1=1$</td>
<td>$C_2=\frac{1}{2}$</td>
<td>$$$3.50</td>
</tr>
</tbody>
</table>

**Solutions:**

**Primal**

$x_1=2$, $x_2=3$

$10x_1+5x_2=\$$3.50$

**Dual**

$w_1=\$$.25$, $w_2=\$$.05$

$12w_1+10w=\$$3.50$. 
Maximize: \( z = 1x_1 + \frac{1}{2}x_2 \)

Subject to:
\[
\begin{align*}
3x_1 + 2x_2 &\leq 12 \\
5x_1 &\leq 10 \\
x_1, x_2 &> 0
\end{align*}
\]

(1) Set up problem in standard form by removing all inequalities

\( E_0: \quad z - x_1 - \frac{1}{2}x_2 = 0 \)

\( E_2: \quad 3x_1 + 2x_2 + x_3 = 12 \)

\( E_3: \quad 5x_1 + x_4 = 10 \)

**Criterion:** Are coefficients of the \( E_0 \) neg.? Yes.

(2) Determine BFS BV's = \( \{ x_2, x_4 \} = \{ 12, 10 \} \) NEV's = \( \{ x_1, x_2 \} = \{ 0, 0 \} \)

(3) Determine entering variable by selecting the column with most negative "shadow price", \( z - c \) in objective function.

That is: \( x_1 \)

(4) Determine leaving variable by selecting the row that becomes the "bottleneck" restriction as the entering variable is increased

\[
\begin{align*}
x_3 &= 12 - 3x_1 - 2x_2 \\
x_4 &= 10 - 5x_1
\end{align*}
\]

\( \begin{array}{c}
\text{Increase in } x_1 \\
4
\end{array} \)

\( \begin{array}{c}
\text{Increase in } x_3 \\
4
\end{array} \)

\( \begin{array}{c}
\text{Increase in } x_4 \\
2
\end{array} \)

\( \begin{array}{c}
\text{The leaving variable is } x_4
\end{array} \)
(5) Determine the new improved solution by solving equations using the \( \eta \)-\( \lambda \) method.

\[
\begin{align*}
E_0' &= E_0 + \frac{1}{5}E_2: z = -\frac{3}{5}x_2 + \frac{1}{5}x_4 = 2 \\
E_1' &= E_1 - \frac{3}{5}E_2: 2x_2 + x_3 - \frac{3}{5}x_4 = 6 \\
E_2' &= E_2: 5x_1 + x_4 = 10
\end{align*}
\]

**Criterion:** All \( z-c \geq 0 \) ?, No.

(2) Determine BFS \( \text{Bfs} = \{x_1, x_3\} = \{2, 6\} \) \( \text{NBfs} = \{x_2, x_4\} = \{0, 0\} \)

(3) Determine entering variable
that is: \( x_2 \)

(4) Determine leaving variable
\[ x_3 = 6 + 3/5x_4 - 2x_2 \]
\[ x_1 = 2 - x_4/5 \]
**: The leaving variable is \( x_3 \).**

(5) Determine the new improved solution by solving equations

\[
\begin{align*}
E_0'' &= E + \frac{1}{4}E_1': z = \frac{1}{4}x_3 + 1/20x = 3.50 \\
E_1'' &= E_1': 2x_2 + x_3 - 3/5x_4 = 6 \\
E_2'' &= E_2: 5x_1 + x_4 = 10
\end{align*}
\]

**Criterion:** All \( z-c \) non-negative? Yes.

**: optimum solution is reached**

\[ x_2 = 2 \ , \ x_3 = 3 \]
\[ z = 3.50 \]
Graphical solution

\[ \text{Minimize: } z' = 12w_1 + 10w_2 \]

subject to:

\[ 3w_1 + 5w_2 \geq 1 \]
\[ 2w_1 \geq \frac{1}{2} \]
\[ w_1, w_2 \geq 0 \]

This problem is equivalent to the following standard form (max and \( \xi \))

Maximize \(-z' = \text{minimize } z'\)

\[ \begin{align*}
  z' + 12w_1 + 10w_2 &= 0 \\
  -3w_1 - 5w_2 &\leq -1 \\
  -2w_1 &\leq \frac{1}{3}
\end{align*} \]
(1) Introduce slack variables to convert inequalities to equalities and artificial variables to eliminate the non-positive $b_i^j(s)$.

\[
\begin{align*}
&z + 12w_1 + 10w_2 = 0 \\
E_1: &3w_1 + 5w_2 - w_3 + w_5 = 1 \\
E_2: &2w_1 - w_4 + w_6 = \frac{1}{2}
\end{align*}
\]

(2) Now, the artificial variables have to be driven to zero. Therefore a dummy (or secondary) objective function is introduced

\[
\text{Min } W = \sum \text{(artificial variables)}
\]

\[
\begin{align*}
&\text{Min } W = w_5 + w_6 \text{ or Max } W = -w_5 - w_6 \text{ or } W + w_5 + w_6 = 0
\end{align*}
\]

(3) Phase I

Before we proceed with the simplex we have to eliminate the BV's $w_5, w_6$ from the objective function.

\[
\begin{align*}
E_0: &w + w_5 + w_6 = 0 \\
E_0': &E_0 - E_1: w - 3w_1 - 5w_2 + w_3 + w_6 = -1 \\
E_0'': &E_0 - E_2: w - 5w_1 - 5w_2 + w_3 + w_4 = -1.5
\end{align*}
\]

\[\text{\textquoteleft\textquoteleft The simplex method in phase I can begin with the equivalent form of the L. P model}\]

\[
\begin{align*}
E_0: &w - 5w_1 - 5w_2 + w_3 + w_4 = -1.5 \\
E_1: &3w_1 + 5w_2 - w_3 + w_5 = 1 \\
E_2: &2w_1 - w_4 + w_6 = \frac{1}{2}
\end{align*}
\]
Initial BFS

\[
\begin{align*}
\{ & w_5 = 0, \\
& w_6 = \frac{1}{2} \}
\end{align*}
\]

Step 1 Determine entering variable: \( w_1 \)

Step 2 Determine leaving variable: \( w_6 \)

\[
\begin{align*}
& w_5 = 1 - 3w_1 - 5w_2 + w_3 - w_5 \\
& \quad \text{leaving} \quad -w_6 = \frac{3}{7} - 2w_1 + w_4 - w_6
\end{align*}
\]

Step 3 Determine new BFS

\[
\begin{align*}
& E_0' = E_0 + \frac{5}{2}E_2: w \\
& E_1' = E_1 + \frac{3}{2}E_2: 15/3w_2 - w_3 + 3/2w_4 + w_5 - 3/2w_6 = .25
\end{align*}
\]

Step 4 Inspect \( E_0 \) for optimality—It can be improved

Cycle 2 (2nd iteration)

Step 1: Determine new entering variable: \( w_2 \)

Step 2: Determine new leaving variable: \( w_5 \)

\[
\begin{align*}
& w_5 = .25 - 5w_2 \\
& \quad \text{increase in} \quad w_2 \quad 1/20
\end{align*}
\]

Step 3: Determine new BFS

\[
\begin{align*}
& E''_0 = E_0' + E_1': w \\
& E''_1 = E_1': 5w_2 - w_3 + 3/2w_4 + w_5 - 3/2w_6 = .25
\end{align*}
\]

Step 4: Check for optimality

—Since all coeff. of \( E''_0 \) are non-negative the solution is optimal.

—Furthermore since min (w) = 0 there is a BFS for Phase II (since \( w_5 = 0, w_6 = 0 \)
Phase II

Now we return to our original objective function (we drop the artificial v's)

\[ E_0: \quad Z + 12w_1 + 10w_2 = 0 \]

To evaluate it for optimality we must eliminate all Bv's \((w_1 \text{ and } w_2)\) from \(E_0\).

\[ E' = E_0 + 6E_2: Z + 10w_2 + 6w_4 = +3 \]
\[ E'' = E_0 + 2E_1: Z + 2w_3 + 3w_4 = +3.50 \]

We notice that the objective function has non-negative coefficients, so the solution is optimal:

\[ w_1 = 0.25 \quad w_2 = 0.05 \]
\[ Z' = 3.50 \]

Graphical solution.
The dual variables are of great significance for the model. The dual variables provide evaluators for changes in capacities in the primal problem.

\[ w_1 = \$0.25 \] means that a one hour increase in the value of \( b_1(=12) \) hours of capacity on the first machine will increase profits by \( \$0.25 \).

and \( w_2 = \$0.05 \) means that a one hour alteration in capacity for \( b_2(=10) \) will add \( \$0.05 \) to total profit.

If we compare the two machines the first machine is to be preferred to the second when optimization is the objective.

Solution to the Input-Output Problem. The Input-Output problem given in page 26 can be considerably simplified (Gavin, referred to in page 26, shows all the necessary calculation) and for

\[
\begin{align*}
x_4 &= 20, \\
x_5 &= 42, \\
x_6 &= 6, \\
x_7 &= 84, \\
x_8 &= 6, \quad P = 1,940
\end{align*}
\]

we have

\[
\text{Max } z = x_1,
\]

Subject to the constraints given below.

\[
\begin{align*}
1.4x_1-x_2+1.2x_3+x_4+3x_5+2x_6+x_7 &= 84 \\
-x_3+3x_4+6x_5+2x_6+x_8 &= 6 \\
x_1 - x_2 &+ x_4 - x_5 + x_9 + x_{10} = 20 \\
x_3 + x_6 + x_{11} &= 6 \\
16x_1 + 21x_2 + 15x_3 + 75x_4 + 180x_5 + 8x_6 + x_{12} &= 1940
\end{align*}
\]
Optimum production program is:

- **Automobile production** = $x_1 = 24$ units
- **Tool production** = $x_3 = 6$ units
- **Automobile capacity** = $x_4 = 4$ units
- **Storage of Steel** = $x_7 = 39.2$ units
- **Steel capacity** = $x_{10} = 42$ units
- **People work force** = $x_{12} = 1,166$ units.
CHAPTER IV

DECISION-MAKING UNDER RISK AND UNCERTAINTY

INTRODUCTION

In this chapter the interaction between the linear programming model on the one hand and the decision making under risk and uncertainty on the other is analyzed and some aspects of the introduction of error into a linear programming model are ascertained. The difficulties arising in formulating a risk programming model and the possibilities of deriving a deterministic equivalent problem are analyzed. In a sense this chapter unifies different viewpoints underlining the foundation of linear programming under risk and uncertainty.

According to Knight there exists an operational distinction between the concepts of risk and uncertainty, both of which describe a different degree of lack of knowledge. His approach issues from the frequency approach to probability. He defines risk by a model which is characterized

*Van Moeseke, Paul, "Stochastic Linear Programming" Yale Economic Essays, Vol. 5., No. 1. (Spring 1965) pp.197-199*
by the knowledge of the entire probability distribution of
the different outcomes, whether the data are subjective or
objective; uncertainty by a model which is characterized by the
lack of knowledge about the probability distribution
of its different outcomes. Basically, under his definition
of uncertainty, the decision-maker is forced, in effect, to
gamble. Both of these restrictive definitions have been
relaxed through the years and in some cases they are used
interchangeably. In linear programming, these terms are
used in a specialized manner and they usually modify the word
programming.

I. ELEMENTS OF UTILITY THEORY

A decision is an evaluation of the merits of a
number of alternatives, A, B, C and the selection of one of
them. According to this theory a decision-maker behaves
as though he places some numerical utility on each certain
outcome and then acts with the objective of choosing the
probability distribution which exhibits the maximum mathematical
expected utility. Decision-making is underlined by two
assumptions. One deals with comparability (the decision

26 Van Moeseke, Paul, "Stochastic Linear Programming"
Yale Economic Essays, Vol., 5, No. 1. (Spring 1965) pp. 197-199

27 Mr. R. W. Morell gives some lexicographic history of
the word and concludes that the word "choice" could be used
synonymously with decision. See: Morell, W. Robert, Managerial
maker prefers either A to B or B to A or he is indifferent between them) and the other with transitivity (If A is preferred to B and B to C, then A is preferred to C as well; the same happens for indifference). When these two assumptions are satisfied the decision maker is able to rank these alternatives according to decreasing preference. This implies that a preference or utility function $F$ exists such that $F(A) > F(B)$ when A is preferred to B and $F(A) = F(B)$ when there is indifference. Then the decision maker's objective may be described by saying that he maximizes $F$ over the set of alternatives available to him.

Usually the decision-maker under conditions of uncertainty selects a probability distribution from a given set of such distributions. Any rational decision-maker will, by mere definition of rationality, select the "best" one from the given distributions. Therefore, his preference orderings will be our sets of probability distributions:

$$f_1(x), \ldots, f_n(x).$$

where:

$$\sum_{i} f_n(x_i) = 1 \quad \text{for all } n.$$

For any given $x$, $f(x)$ signifies the probability that $x$ will occur and is called the prospect $p$ of the set of

elements of prospects.

\[ f(x_0), f(x_1), f(x_2), \ldots, f(x_p) \ldots \]

The preference ordering over the set of prospects can be represented by a utility functional \( u(f) \) in the following manner:

\[ u(f_1) > u(f_j) \Rightarrow f_1 > f_j \]

and \( u(f) \) can be written

\[ u(f) = u(\ldots, f(x_0), f(x_1), f(x_2), \ldots, f(x_p); \ldots) \]

There are three axioms which govern the ordering method:

**Axiom 1:** To any prospect \( f(x) \) in the set there corresponds a certainty equivalent \( \bar{x}, (1, \bar{x}) \sim f(x) \).

**Axiom 2:** If \( x \) is defined over the domain \( 0 \leq x \leq g \) and if the assumption is that the only two possible outcomes are:

- \( g \) with probability \( p \)
- \( 0 \) with probability \( 1 - p \)

Then the prospect is binary and is given by \( (p, g) \). Now, from Axiom 1 all such prospects have certainty equivalents; i.e. for any \( p \), there exists a number \( x_p \), such that

\( (1, x_p) \sim (p, g) \).

As \( p \) increases from 0 to 1, \( x \) will be increased from 0 to \( g \).

The best description of a prospect can be given by stating that the awards \( 0, 1, \ldots, r, \ldots, g \) can be won with the probabilities \( f(0), \ldots, f(r), \ldots, f(g) \) respectively.

Now, if the award \( r \) is replaced by its equivalent binary prospect \( (p_r, g) \), it will give a modified prospect \( f(r)(x) \),
where there is no award equal to \( r \), then it can be seen that

\[
\begin{align*}
 f^r(0) &= f(0) + f(r)(1 - p) \\
 f^r(1) &= f(1) \\
 f^r(2) &= f(2) \\
 & \vdots \\
 f^r(r) &= 0 \\
 f^r(g) &= f(g) + f(r)p_r
\end{align*}
\]

A rational decision-maker will be indifferent if his prospect is modified in this way or not, since \((1, r, p_r, g)\). Therefore the following axiom can be stated:

**Axiom 3**: \( f(x) \) and \( f^r(x) \) have the same certainty equivalent.

Now if this axiom is applied to all the awards except the boundary \( o \) and \( g \), then the result will be a prospect of the type \((p, g)\), which has the same certainty equivalent as the original prospect \( f(x) \). \( P \) is given by:

\[
p = p_1 f(1) + p_2 f(2) + \ldots + p_{g-1} f(g-1) + f(g)
\]

This is true because of Axiom 2 according to which \( p_0 = 0 \) and \( p_g = 1 \). Therefore, \( p \) can be written

\[
p = \sum_{x=0}^{g} p_x f(x)
\]

From the above function a complete preference ordering over the set of prospects \( f(x) \) can be obtained. For any given arbitrary prospects \( d(x) \) and \( h(x) \) the corresponding prospects \( (p_d, g) \) and \( (p_h, g) \) and their certainty equivalents can be computed; and \( d(x) \) is preferred to \( h(x) \) if and only if \( p_d > p_h \). Then a utility function can be defined over the preference...
ordering as
\[ u(f(x)) = p_d = \sum_{x=0}^{M} p_x f(x) \]
\[ \text{or} \]
\[ u(f(x)) = \sum_{x=0}^{M} u(x)f(x) \text{ where: } p_x = u(x). \]
and this formula is called the Expected Utility Hypothesis.

II. CHARACTERIZATION OF DECISION PROCESS

Decision Processes have a vertical escalation and a horizontal classification. Vertically the decision processes fall into the following categories:

1. **Strategic decisions:** These are either long term decisions saddled with multiple objectives or short-term policy decisions pertaining to one objective.  

2. **Tactical decisions:** These are the means by which strategy is implemented for achievement of the objective.  

3. **Operational decisions:** These are the detailed day by day period by period decisions which implement the decided tactics of the organization.

Horizontally the classification of decision processes depends on the amount of information available for assessment of the "state of things" and decision-making. The

---


chart below shows the information spectrum and the relative classification of decision-making.

Information with known probabilities
Full information of occurrence No information

Decision processes under certainty Decision processes under risk Decisions processes under uncertainty

In Operation Research parlance the three divisions are usually referred to as: deterministic, stochastic and decision-making under uncertainty. Of course, in reality there is no clear-cut division. Problems tend to appear at any point in the information spectrum, but the rational approach to decision-making recognizes the necessity of approximation and therefore the decision-making problems could be truncated to the closest one of the three main points.

In engineering, the decision-making processes follow the horizontal classification but fail to identify with the vertical one. The engineers in decision theory emphasize more the nature of the data from which they have to make decisions, and consequently they have the following classifications.

For an extensive coverage of these topics see: Bierman et. al., Quantitative Analysis for Business Decisions, (Richard D. Irwin, Inc., Homewood, Illinois, 1955)

1. A single selection is made from a finite number of alternatives (presence of discreet data).

2. A single value is chosen from a continuous range of possible values (presence of continuous data).

3. Both the above operations are performed.

The decision processes are usually characterized by high sensitivity; that is, vulnerability to small changes in the controlling variables, conflicts between total system objective optimization and suboptimization; that is, as attempt to optimize tactical decisions with little or no regard for the overall strategic decision situation, and the presence of intangible factors which are not quantified for inclusion in the decision model, i.e. public relations, personnel gratification etc. Further, the decision processes are obstructed by problems of data collection, lack of classificatory systems and lack of sequential system control.

In order to delineate the problems of decision-making and expedite solution, abstractions from the full detail of reality are made to reason out, in advance, the contingencies that might be encountered and thereby secure an opportunity for choosing rational decision rules to subsequent actions.\(^{34}\) The decision-making process steps through the following procedures.

\(^{34}\)Riggs, L. J. op. cit. p. 13.
1. Recognition of the existence of a problem.
2. Definition of the problem.
3. Collection of all relevant informational and structural data.
4. Formulation of an appropriate model.
5. Simulation.
6. Evaluation of the model.
7. Decision-making through the model.
8. Decision implementation and acceptance of feedback for modification of the model.35

The quantitative methods which engulf the above procedure are mainly statistical and operations research oriented. They fall into three classes.

2. Bayesian statistics.
3. Operations research.

The first two classes of the quantitative methods are mainly employed in the micro-operational decision area. Operations research spans the two fields of "micro" and "macro" operations within the operations research field; linear programming is the most advanced technique in reference to applications for decision-making.

35 Alternative sets of procedures are given in some parts of the literature. See: Folsom, E., Marion, Executive Decision Making, (McGraw-Hill Book Co., Inc., New York, 1962) p. 4
III. SOURCES OF UNCERTAINTY IN THE MODEL DATA

Uncertainty impinges on the model from errors in the data base which are expressions of either model imperfection and incompleteness of description or numerical failures in establishing the data base. Oskar Morgenstern\footnote{Morgenstern, Oskar, \textit{On the Accuracy of Economic Observations}, (Princeton University Press, Princeton, New Jersey, 1963) pp. 13-61.} gives the following main sources of errors due to model imperfection and incompleteness of description.

1. **Lack of designed experiments:** Most of the data for mathematical models are derived from experiments that their main emphasis is not the preservation of accuracy as much as the approximate representation of a real-life system.

2. **Information hoarding:** Data accumulated from various statistical sources which depend on their accuracy on the openness and truthfulness of a human source suffer from evasive answers and deliberate lies due to the fact that the human source of information suspects that in some way data will be used against him.

3. **Untrained data collectors:** The accumulation of data is sometimes the responsibility of individuals which are not scientifically trained. Even trained personnel most of the time bias the data by their preference for certain sort of data and their dislike for others.
4. **Errors from questionnaires:** There are two obstacles to the method of questionnaires for data gathering. One is the preparation of the questionnaires in a manner that will insure a sound data basis for the objective of the model, and second the answering of the questionnaire depends very much on the emotional prejudices of the person filling the form.

5. **Mass observations:** In big organizations a large mass of data is accumulated and usually it is aggregated through various statistical processes. The successive processing of data may cumulate the errors with no guarantee of cancellation.

6. **Lack of classification or misclassification:** Due to improper or incomplete classification of the data, the whole data basis is subject to serious introduction of errors especially when the data is aggragated.

7. **Errors of instruments:** Instruments are usually calibrated for accuracy within certain intervals. If the original data readings have already the element of error in them, then processing compounds the error without the benefit of cancellation.

8. **The factor of time:** If data is accumulated at different intervals of time with variable interval length then the process of interpolation necessary to complete the data basis brings its own seed of an error aggravated situation.

The above classifications cover more or less in general the sources of error that can be introduced into a mathematical model before even the actual computation takes place. Every
mathematical model carries its own built-in tendency to negate its purpose due to the above mentioned general sources and the sources of error peculiar to the individual models. The linear programming model does not escape this tendency.

IV. ERRORS AND STABILITY IN LINEAR PROGRAMMING

The basic assumption underlying the linear programming theory is the assumption of perfect knowledge about the parameters of the linear programming model. As it was seen in the previous section this could prove to be an unrealistic assumption. The linear programming model has been used in diverse applications with considerable success, and has effected enormous savings in economic planning, it is nevertheless true that proportionally only a very small higher percentage of savings was achieved over more conventional forms of planning and optimization. Very small errors or fluctuations in the data could very easily eliminate the relative advantage of linear programming over cruder methods. This fact can be demonstrated by the following linear programming problem:\textsuperscript{37}

\[ \text{Maximize } z = 15x_1 + 30x_2 - 2x_4 + 50x_5 \]

subject to

\[ \begin{align*}
3x_1 + 7x_2 + x_3 - 3x_4 &= 10 \\
-20x_1 - 2x_2 - 6x_3 + 21x_4 + 3x_5 &= 13 \\
4x_1 + 17x_2 + x_3 - 4x_4 + 11x_5 &= 7
\end{align*} \]

The solution of this linear program is:

\[ \begin{align*}
x_1 &= 64, x_2 = 0, x_3 = 19, x_4 = 57, x_5 = 0
\end{align*} \]

and the required maximum value is 1525.

Now, if each coefficient \( a_{ij} \) is liable to error with a coefficient of variation of 2\% and the linear program is solved using expected values for the parameters then the solution will be as follows.

\[ \begin{align*}
E(x_1) &= 97, E(x_2) = 20, E(x_3) = 100, E(x_4) = 0
\end{align*} \]

and the expected maximum value \( E(z) \) is increased to 2315.

It is noted that a 2\% variation in each of the coefficients produces considerable changes in the basic variables and an increase of nearly 50\% in the maximum value. The fact is of course that the stability of the solution depends very much on the structure of the matrix of the coefficients. There is no reason, however, in practice, to suppose that a linear program must be better conditioned or worse conditioned that the example above.

The standard deviations of the basic variables were computed to be

\[ \begin{align*}
s(x_1) &= 47 \\
s(x_2) &= 31 \\
s(x_3) &= 47
\end{align*} \]

This example brings into focus the necessity for analysis
of the sensitivity of an optimal solution with changes in
the parameters of the model.

Another reason for study of the sensitivity of the
solution is that frequently in practical applications the vectors
b or c of the linear programming program are not constant
but random variables and the effects of specific assumptions
about the probability distribution of the elements of the
model in setting confidence intervals for the optimal solution
have to be studied. In addition, there is a need for studying
the appropriate safety margins when small violations of the
constraints are allowed.

All the above reasons make it imperative for risk to be
introduced into the linear programming model.

The introduction of risk into the model brings the
utility functional and the formal statistical decision theory
in direct relationship with linear programming.

Many thumb-rule methods were devised in the past to take
care of stochasticity in the elements of the linear programming
models, especially in applications involving scheduling. The
standard technique for the reduction of the effect of probabilistic
or chance events has been the providing of plenty of fat in the
system. For example, wear-out rates, consumption rates,
attrition rates have been planned on the high side, always
above the actual data with the hope that they would be used
as "shock absorbers" which would permit the system to operate
in spite of unforeseen events.
The concept of mathematical expectation was also introduced, especially in the objective function of the model, where the randomization of the coefficient vector $\bar{c}$ made the optimization of $\bar{z}$ meaningless. This is a tempting short-cut method which drops the probability alternatives of utility altogether and puts the random deviations at their expectation values or at their most probable values.\footnote{Naslund, B., and A. Whinston, "A Model of Multi-period Investment Under Uncertainty", Management Science, Vol.8 (1962) p. 184.} This short-cut is permissable, in the sense that decision maker assumes that the law of large numbers is in operation and all the information he needs concentrates in the expected value of distribution. Most of the probabilistic models developed, consistently employ the expected value of the objective function as the goal criterion.\footnote{Madansky A. "Linear Programming Under Uncertainty" in Recent Advances in Linear Programming. Wolfe (ed.),1960, pp. 103-110.} Moreover, if $Z$ is a linear function and the objective of the model is to optimize (min. or max.) the expected value of $Z,E(Z)$, then $E(Z)$ is a function of the random vector $\bar{c}$ and the decision variables $x$ is not necessarily linear in $x$, and this gives rise to non-linear deterministic programming problems.\footnote{Freund, R.J. "The Introduction of Risk into a Programming Model", Econometrica, Vol. 24 (1956) p. 255.} There are many other problems associated with the solution of stochastic programming problems and some of them will be illustrated in the later chapters.
CHAPTER V

STOCHASTIC LINEAR PROGRAMMING

INTRODUCTION

A linear programming problem is said to be stochastic, if one or more of the coefficients in the objective function or the system of constraints or availabilities is known only by its probability distribution. The first who suggested the name and an appropriate method of solution for a subset of the general stochastic problem was Tintner. Babbar, with the help of Tintner, did some fundamental work on the distributions of the solution of a set of linear equations, when the coefficients are subject to random errors.

A distinction is generally drawn between two different approaches to stochastic linear programming: The passive (also termed "wait and see" approach) and the active (also termed "here and now" approach). In this chapter the two approaches will be presented and some theorems pertaining to


the dual problem of stochastic linear programming will be given. Lately, several results were found connecting the two approaches of stochastic linear programming. These results extend the theoretical foundations of stochastic linear programming but they have no applicability to real-world problems.

I. PASSIVE STOCHASTIC PROGRAMMING

Let $A = (a_{ij})$ be an $mxn$ matrix, let

$$\bar{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$$

be an $n$-dimensional column vector, let

$$\bar{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_m \end{bmatrix}$$

be an $m$-dimensional column vector and let

$$\bar{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_n \end{bmatrix}$$

be an $n$-dimensional column vector. Then the deterministic linear programming model can be stated as

$$\text{maximize } z = \bar{c} \bar{x} \quad (5.1)$$

subject to
\[ A \bar{x} \preceq \bar{b} \quad (5.2) \]

and

\[ \bar{x} \succ 0 \quad (5.3) \]

where the accent indicates transposition of the vector. The inequalities (5.2) can be transformed into equalities according to the results of Chapter III. Let the vector of the slack variables be denoted by \( \bar{s} = I s_1, \ldots, s_m I \), then the new vector is formed denoted by

\[ \bar{y} = (\bar{x}, \bar{s}) \quad (5.4) \]

and the new matrix of the coefficients

\[ B = [A, I] \quad (5.5) \]

where \( I \) is the unit matrix of order \( m \). Also, in order to carry over the variable transformation to the objective function (5.1) a new vector is formed having \( nt \) elements

\[ \bar{d} = (\bar{c}, \bar{0}) \quad (5.6) \]

where \( \bar{0} \) is a vector having \( m \) zero components. Now, the maximization problem can be stated as follows:

Maximize \( G = \bar{d} \cdot \bar{y} \quad (5.7) \)

subject to

\[ B \bar{y} = \bar{b} \quad (5.8) \]

and

\[ \bar{y} \succ 0 \quad (5.9) \]

The problem has to be solved for the variables \( \bar{y} \), and numerically this can be done very efficiently with the simplex method of Chapter III. However, for purposes of the analysis another method will be used, which though it is less rapid in
giving the solution, illustrates better the stochastic problem. The method is called the method of selections\textsuperscript{43}. 

Let $r$ be the rank of the matrix $A$. Then $r$ elements of the vectors $\overline{y}$ and $\overline{d}$ and $r$ columns corresponding to the matrix $B$ can be selected. This selection has to be affected by all the possible combinations. The total number of selections is $K = \binom{n}{r} = \frac{n!}{(n-r)!r!}$. In other words the number of selections is the number of combinations of $n$ elements taken $r$ at a time. Let $\overline{y}(k)$ be a fixed selection of the elements of the vector $\overline{y}$, $\overline{d}(k)$ the corresponding choice of the elements of the vector $\overline{d}$ and $B(k)$ the corresponding selection of the columns of the matrix $B$. Then $r$ linear equations can be obtained from the elements of the vector $\overline{y}(k)$:

$$B(k) \cdot \overline{y}(k) = \overline{b} \quad (5.10)$$

The solution of the system (5.10) can be obtained by inverting the matrix $B(k)$. Therefore:

$$\overline{y}(k) = (B(k))^{-1} \cdot \overline{b} \quad (5.11)$$

Now, the solutions given by (5.11) will constitute a feasible set if all the components of the vector $\overline{y}(k)$ are non-negative\textsuperscript{44}. Therefore for all feasible solutions the

\textsuperscript{43}Dorfman, R., Applications of Linear Programming to the Theory of the Firm, (Berkeley, California, 1951) p. 31.

following shows the effect on the objective function
\[ G(k) = \left( \overrightarrow{d(k)} \right)' \cdot \overrightarrow{y}(k) \] (5.12)
which is the net gain for the selection K. From this set of feasible solution \( G(k) \)'s the greatest can be found. This greatest \( G(k) \) is the solution to the maximization problem and is denoted by
\[ G^* = \max G(k) \quad (k=1, \ldots, k) \] (5.13)

Up to the present time the formulation has been deterministic. Now the concept of risk can be introduced into the model by assuming that all the elements of \( \bar{c}, \bar{b} \) and \( A \) are random variables subject to probability distributions. A further assumption is that they are independently distributed with probability density functions given by, say
\[ \int_{a_{ij}} a_{ij}, \int_{b_i} b_i \text{ and } \int_{c_j} c_j \] (5.12)
Hence there variables have a joint probability density function given by
\[ P(a_{ij}, \ldots, a_{mn}, b_i, \ldots, b_m, c_i, \ldots, c_n) = \]
\[ = \prod_{i=1}^{m} \prod_{j=1}^{n} a_{ij} b_i c_j \] (5.13)
or
\[ P(a_{11}, \ldots, a_{mn}, b_1, \ldots, b_m, c_1, \ldots, c_n; w_1, \ldots, w_t) \]

having as domain the product parameter space
\[ S = \prod_{i=1}^{m+n} R_i \] (5.14)
Where $R_i$ represents the real line as defined in Chapter II, and the $W_t$ are the parameters fixing the form of the function. Therefore, the totality of the elements of $A$, $b$, $c$ and of the possible variations given by the probability function (5.13) are found within $S$. In a deterministic linear program the parametric space $S$ is reduced to a single point due to the definitions given in Chapter II for linear spaces. The feasibility condition of $\bar{y}(k) \geq 0$ ($k=1, \ldots, k$) will give a region $S_k$ in $S$, in which this condition is satisfied. The set of regions $S_1, S_2, \ldots, S_k$ is called the set of feasible regions and by overlapping of its elements a system of subregions in $S$ is formed.

Let $T_k$ be the set in $S$ for which condition (5.13) is satisfied. This is the region in which the selection $k$ is greatest than the other selections and the objective function attains its maximum. But $G(k)$ is not necessarily feasible in the region $T_k$. Therefore for the solution to be feasible and optimum it has to belong to both sets $S_k$ and $T_k$. The introduction of the set $U_k = S_k \cap T_k$ allows the optimality and feasibility conditions to be met. The set $U_k$ being the intersection of the sets $S_k$ and $T_k$ is characterized by the following properties: in $U_k$, the selection $k$ is possible ($\bar{y}(k) \geq 0$) and also the objective function $G(k)$ is maximum. Now, if $G^{**}$ is the best linear function at a given point in $S$ and if the elements of the vectors $b^{(k)}$ and $c^{(k)}$ and of the matrix $A^{(k)}$ are part of the set $U_k$ then $G^{**} = G(k)$. 


G** then is a function of the elements of the vectors and of the matrix

\[ G^{**} = G(k) (A(k), \bar{b}(k), \bar{c}(k)) \]  

(5.15)

But the elements of the vectors and of the matrix are random variables with a joint probability density function given by (5.13) and G** is a function of these random variables according to (5.15), therefore G** itself is a random variable.

It is possible to derive the probability distribution of G** either through analytical means or through numerical simulation means. Suppose that a set of variates drawn from the probability distributions for the random elements in A, \( \bar{b} \) and \( \bar{c} \) is substituted for stochastic parameters in the problem. If the resulting nonstochastic components are denoted by A, \( \bar{b} \) and \( \bar{c} \) then the value of the objective function in the solution to the associated deterministic linear programming problem G is a variate from the probability distribution \( \varrho_2 \) \((G^{**}; w_1, w_2, \ldots, w_t)\) of G**. By repeating this procedure through Monte Carlo randomization methods the distribution \( \varrho_2 \) can be approximated. Now it becomes clear why this is called the "wait and see" or passive approach of stochastic linear programming.

This numerical method used in approximating \( \varrho_2 \) \((G^{**}; w_1, w_2, \ldots, w_t)\) involves the solution of a number of deterministic problems.

linear programming problems, with observed values substituted for the random variables. The sets of random variables are supposed to be obtained from any given or assumed probability distributions. If the probability distributions associated with the risky coefficients in the stochastic program are normal, the methods for obtaining the necessary sets of random variates is simple and well-known. Recently a paper was published giving an algorithm for generation of random variates from any given probability distribution. This algorithm is extremely efficient in connection with a fast algorithm for the solution of the general deterministic linear programming given in Llewellyn. These two algorithms together could give the distribution of the objective function of any linear stochastic problem to any desired accuracy, subject to the computer precision capability and the degree of interpolation and the number of points used for estimating the transformation function of the random deviates according to the Theory given by Mood.


52 This algorithm is based on the MINIT method given by Llewellyn, R., W. Linear Programming (Holt, Rinehart and Winston, New York, August 1966) pp. 207-218.

The derivation of the probability function of $G^x$ analytically is a very difficult problem\textsuperscript{54}. Let

$$M \cdot \bar{w} = \bar{q} \quad (5.16)$$

be a system of linear equations. The examination of small perturbations around the values of the elements of $M$, $\bar{w}$ and $\bar{q}$ will be done by taking the derivatives of the matrix $M$ and of the vectors $\bar{w}$ and $\bar{q}$.

$$dM \cdot \bar{w} + M \cdot dw = dq \quad (5.17)$$

Then

$$dw = -M^{-1} \cdot dM \cdot \bar{w} + M^{-1} \cdot dq \quad (5.18)$$

where $dw$ is a linear approximation for small perturbations of the solution vector $\bar{w}$. Now, since the elements of $A$, $\bar{b}$, and $\bar{c}$ in (5.1) and (5.2) are random variables, their variations being the result of many small deviations are assumed to tend towards the Laplace-Gauss function (normal probability function). This assumption is mainly justified by the Law of large numbers\textsuperscript{55}.

With the use of formulas (5.18) and (5.11) the deviation of the activity vector $\bar{y}(k)$ from its mean value can be given as

$$dy(k) = (B_o(k))^{-1} \cdot dB - (B_o(k))(dB(k))\bar{y}_o(k) \quad (5.19)$$


where: $\overline{y}_o(k)$ = vector of arithmetic means.
$B_o(k)$ = matrix of arithmetic means.

Formula (5.19) can be reduced to
$$\frac{\overline{y}_o(k)}{\overline{y}_o} = (B_o(k)^{-1} \cdot B_o$$

By using formulas (5.18) and (5.12) the following result appears
$$dG(k) = (d\overline{y}(k)) \cdot \overline{y}_o(k) + (d_o(k)) \cdot d\overline{y}(k)$$
(5.21)

and since (5.21) is the deviation of the objectione function from its arithmetic mean
$$G_o(k) = (d_o(k)) \cdot \overline{y}_o(k)$$
(5.22)

Now, with the help of the given formulas the approximations
$$\overline{y}(k) = \overline{y}_o(k) + d\overline{y}(k)$$
and
$$G(k) = G_o(k) + dG(k)$$
can be easily calculated for all the points of the parametric space $S$ and their probabilities can be determined.

As an example of a situation when knowledge of the distribution of max. $Z$ is of value to a decision-maker consider a farmer who has to decide the most profitable way of distribution of his resources in planning his farm production 56. Only two activities are considered: The growing of corn ($X_1$) and the growing of flax ($X_2$).

The input coefficient of land for corn is $a_{11}$.
The input coefficient of land for flax is $a_{12}$.
The input coefficient of capital for corn is $a_{21}$.
The input coefficient of capital for flax is $a_{22}$.

The amount of land available is \( b_1 \).

The amount of capital available is \( b_2 \).

The means and standard deviations for \( a_{ij} \) (\( i=1,2; j=1,2 \)) are known. The \( b_i \) (\( i=1,2 \)) are constant. Also it is assumed that the net prices of corn and flax are \( c_1 \) and \( c_2 \) and they are constant.

The linear program can be formulated as follows.

Maximize \( Z = c_1x_1 + c_2x_2 \)

Subject to

\[
\begin{align*}
\begin{array}{c}
\end{align*}
\end{align*}
\]

\[
\begin{align*}
\begin{array}{c}
\end{align*}
\end{align*}
\]

\[
\begin{align*}
\begin{array}{c}
\end{align*}
\end{align*}
\]

\[
\begin{align*}
\begin{array}{c}
\end{align*}
\end{align*}
\]

or \( Bz = b \)

where

\[
B = \begin{bmatrix}
a_{11} & a_{12} & 1 & 0 \\
a_{21} & a_{22} & 0 & 1 \\
\end{bmatrix}
\]

and

\[
\bar{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}
\]
The rank of B is 2 and therefore \( \binom{4}{2} = 6 \).

In other words there are 6 ways that two columns of B and the corresponding two elements of \( \bar{z} \) can be selected.

1st selection

\[
\begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\
b_2
\end{pmatrix}
\]

\[
y_1(1) = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11} a_{22} - a_{21} a_{12}}
\]

\[
y_2(1) = \frac{a_{11} b_2 - a_{21} b_1}{a_{11} a_{22} - a_{21} a_{12}}
\]

\[
z(1) = c_1 y_1(1) + c_2 y_2(1)
\]

2nd selection

\[
\begin{pmatrix}
a_{11} & 1 \\
a_{21} & 0
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_3
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\
b_2
\end{pmatrix}
\]

\[
y_1(2) = + \frac{b_2}{a_{21}}
\]

\[
y_3(2) = + \frac{-a_{11} b_2 + a_{21} b_1}{a_{21}}
\]

\[
z(2) = c_1 y_1(2) + c_2 y_3(2)
\]

3rd selection

\[
\begin{pmatrix}
a_{11} & 0 \\
a_{21} & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_4
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\
b_2
\end{pmatrix}
\]
\[ y_1(3) = \frac{b_1}{a_{11}} \quad (5.29) \]
\[ y_4(3) = \frac{a_{11} b_2 - a_{21} b_1}{a_{11}} \]
\[ z(3) = c_1 y_1(3) + c_2 y_4(3) \]

4th selection

\[
\begin{pmatrix} a_{12} & 1 \\ a_{22} & 0 \end{pmatrix} \begin{pmatrix} y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}
\]

\[ y_2(4) = \frac{-b_2}{a_{22}} \quad (5.30) \]
\[ y_3(4) = \frac{a_{12} b_2 - a_{22} b_1}{a_{22}} \]
\[ z(4) = c_1 y_2(4) + c_2 y_3(4) \]

5th selection

\[
\begin{pmatrix} a_{12} & 0 \\ a_{22} & 1 \end{pmatrix} \begin{pmatrix} y_2 \\ y_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}
\]

\[ y_4(5) = \frac{b_1}{a_{12}} \]
\[ y_4(5) = \frac{a_{12} b_2 - a_{22} b_1}{a_{12}} \quad (5.31) \]
\[ z(5) = c_1 y_2(5) + c_2 y_4(5) \]

6th selection

\[
\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \]
\[ y_3(6) = \frac{b_1}{1} = b_1 \]
\[ y_4(6) = \frac{b_2}{1} = b_2 \]
\[ z(6) = c_1y_3(6) + c_2y_4(6) \]

These selections are feasible if \( Y_i(k) \geq 0, (i=1, \ldots, 4; k=1, \ldots, 6) \). The points defined by \( Z(k) (k=1, \ldots, 6) \) constitute a distribution of the optima, and the farmer can decide according to the prevailing values of the \( a_{ij} \) what his optimum policy should be.

The passive approach is intended for the estimation either exactly or approximately of the cumulative statistical distribution of the optimum of a linear programming problem when the joint probability density of the random coefficients is known. Therefore the main applications of this approach lie in testing the sensitivity of the optimum of a stochastic linear program and in developing alternative deterministic equivalents.
II ACTIVE STOCHASTIC PROGRAMMING

Tintner formulated what he calls the "active approach" to stochastic linear programming in 1960. He defines the active approach by introducing new decision variables $u_{ij}$, which are allocation ratios $0 \leq u_{ij} \leq 1$ such that

$$b_i = \sum_{j=1}^{n} b_{ij} u_{ij}; \quad i=1, \ldots, m.$$ \hspace{1cm} (5.33)

Under this approach the linear programming model becomes:

Maximize \hspace{1cm} $z = \bar{c} \times \bar{x}$ \hspace{1cm} (5.34)

Subject to

$$AX = BU \hspace{1cm} (5.35)$$

$$0 \leq u_{ij} \leq 1$$

Where $U = [u_{ij}]$ denotes that resource $i$ which is allocated to activity $j$. $X$ and $B$ are an $n$-order and $m$-order matrices respectively where the diagonal elements of $X$ and $B$ are the elements of vector $\bar{x}$ and $\bar{b}$ respectively.

One of the specific advantages of the active approach is that it allows a flexible formulation of the decision-making problem. For instance if the system of formulas (5.34) and (5.35) is interpreted as an economic policy model, then the activities vector $\bar{x}$ can be regarded as the target vector,

while the elements $U_{ij}$ of the matrix $U$ may be regarded as the set of instrument variables.\footnote{Theil, H. "Econometric Models and Welfare Maximisation" Weltwirtschaftliches Archiv., vol. 72, (1959) pp. 60-83.}

The main assumption in the model is that all resources are completely used. Then, if the probability distribution $P(\tilde{c}, A, \tilde{b})$ is assumed to be known, the probability distribution of the anticipated net profits $Z$ can be derived by numerical randomization techniques or analytically. The probability distribution of $Z$ can be denoted by

$$R(Z; U)$$

which means that it depends on the choice of the variables $U_{ij}$. Then, a preference functional can be defined and the maximization can take place over this preference functional with respect to the elements of the matrix $U$. A special case in the model could be developed if instead of trying to find the distribution of $Z$, only its expected value is wanted. For this case the objective function becomes

$$EZ = \int Z \cdot dR(Z; U)$$

where the integral is to be taken over the whole range of variation of $Z$. Bellman devised a method using dynamic programming for maximizing the mathematical expectation of an integral function.\footnote{Bellman, R., Dynamic Programming (Princeton University Press, Princeton, 1957) pp. 61-80 and 222-244.}
Now, going back to the farmer's example, the problem can be formulated in the active approach as follows:

Maximize \( Z = C_1X_1 + C_2X_2 \)

Subject to

\[
\begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
\leq
\begin{pmatrix}
b_1 & 0 \\
0 & b_2
\end{pmatrix}
\begin{pmatrix}
u_{11} & u_{12} \\
u_{21} & u_{22}
\end{pmatrix}
\quad (5.39)
\]

OR

\[
a_{11}x_1 \leq b_1u_{11}
\]
\[
a_{12}x_2 \leq b_1u_{12}
\]
\[
a_{21}x_1 \leq b_2u_{21}
\]
\[
a_{22}x_2 \leq b_2u_{22}
\]

and then proceed to examine the distribution of \( \text{max } Z \) for various \( U_{ij} \geq 0 \) allocations which satisfy

\[
\begin{aligned}
U_{11} + U_{12} &= 1 \\
U_{21} + U_{22} &= 1
\end{aligned}
\quad (5.41)
\]

The objective is to choose a best set of \( U_{ij} \) values for the matrix \( U \) as judged relative to a suitably defined preference functional.

The preference functional for this example can be defined as

\[
\hat{f} = f \left( R(Z; U_{11}, U_{12}, U_{21}, U_{22}) \right)
\quad (5.42)
\]

which can now be maximized by choice of the proportion of the resources to be denoted to the two activities.

Now, in the example the first thing to be determined is the probability distribution \( R \) under the conditions that
Ul1 is the percentage of b1 devoted to growing corn (activity x1); U12 the proportion of the land devoted to growing flax (x2). In the same fashion, U21 is the proportion of the available capital b2 devoted to corn production, U22 the proportion of capital used in growing flax. The input coefficients a11, a12, a21, a22 are assumed to be normally distributed independent random variables with means and standard deviation given as follows:

<table>
<thead>
<tr>
<th>Input Coefficient</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>E(a11)</td>
<td>S(a11)</td>
</tr>
<tr>
<td>A12</td>
<td>E(a12)</td>
<td>S(a12)</td>
</tr>
<tr>
<td>A21</td>
<td>E(a21)</td>
<td>S(a21)</td>
</tr>
<tr>
<td>A22</td>
<td>E(a22)</td>
<td>S(a22)</td>
</tr>
</tbody>
</table>

The distribution of net profits Z is derived by noting that:

\[ Z = \frac{c_1 b_1}{a_{11}} + \frac{c_2 b_2}{a_{12}} \text{ if: } U_{11}/a_{11}, U_{21}/b_{21}, U_{12}/b_{12}, U_{22}/a_{22} \]

\[ Z = \frac{c_1 b_1}{a_{11}} + \frac{c_2 b_2}{a_{22}} \text{ if: } U_{11}/a_{11}, U_{21}/a_{21}, U_{12}/a_{12}, U_{22}/b_{22} \]

\[ Z = \frac{c_1 b_1}{a_{21}} + \frac{c_2 b_2}{a_{12}} \text{ if: } U_{11}/a_{11}, U_{21}/a_{21}, U_{12}/a_{12}, U_{22}/a_{22} \]

\[ Z = \frac{c_1 b_1}{a_{21}} + \frac{c_2 b_2}{a_{22}} \text{ if: } U_{11}/a_{11}, U_{21}/a_{21}, U_{12}/a_{12}, U_{22}/a_{22} \]

Again the distribution of Z has been derived by numerical methods. The random variables aij have been replaced by corresponding normal independent standardized variables and only the following values have been considered:
0, ±1, ±2, ±3.

Also only the following values have been considered for $U_{ij}$:

$0, 1/4, 1/2, 3/4, 1$.

Then the mathematical expectation of profits can be given in the following Table.

Table V- Expected Values of $Z$

<table>
<thead>
<tr>
<th>$U_{11}$</th>
<th>0</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$Z_{11}$</td>
<td>$Z_{12}$</td>
<td>$Z_{13}$</td>
<td>$Z_{14}$</td>
<td>$Z_{15}$</td>
</tr>
<tr>
<td>.25</td>
<td>$Z_{21}$</td>
<td>$Z_{22}$</td>
<td>$Z_{23}$</td>
<td>$Z_{24}$</td>
<td>$Z_{25}$</td>
</tr>
<tr>
<td>.50</td>
<td>$Z_{31}$</td>
<td>$Z_{32}$</td>
<td>$Z_{33}$</td>
<td>$Z_{34}$</td>
<td>$Z_{35}$</td>
</tr>
<tr>
<td>.75</td>
<td>$Z_{41}$</td>
<td>$Z_{42}$</td>
<td>$Z_{43}$</td>
<td>$Z_{44}$</td>
<td>$Z_{45}$</td>
</tr>
<tr>
<td>1</td>
<td>$Z_{51}$</td>
<td>$Z_{52}$</td>
<td>$Z_{53}$</td>
<td>$Z_{54}$</td>
<td>$Z_{55}$</td>
</tr>
</tbody>
</table>

Now, the $Z_{ij}$ ($i=1,\ldots,5; j=1,\ldots,5$) are the mathematical expectation of profit under different allocations of $U_{ij}$. More precisely the figures in this table are interpreted in the following manner: if the farmer devotes $1/4$ of his land to corn ($U_{11}=0.25$) and hence $3/4$ of the land to flax ($U_{12}=0.75$) and $1/2$ of his capital to corn ($U_{21}=0.5$) and hence also $1/2$ of the capital to flax ($U_{22}=0.5$), then the mathematical expectation of his profit is found by averages the two mathematical expectations $Z_{22}$ and $Z_{33}$. Therefore under the above conditions $E(Z)=(Z_{22}+Z_{33})/2$. Now, if the farmer studies the table of the expected values of $Z$ under different $U_{ij}$, he will be able to choose the best policy for allocating
his land and capital to produce maximum expected profit.
In practice the decision-maker can even specify a confidence
interval under which he will try to maximise the profit.
For instance the farmer can try to maximize Z profit which
he can make with a 95 per cent probability. The numerical
approximations will give again a table like the one above with
analogous interpretation.

The active approach of stochastic linear programming
has been applied to many fields and the results are en­
couraging, although the errors due to approximation formulas
necessitate the invention of better numerical algorithms for
solution of the problem. A very interesting application of
the active approach is to stochastic decision trees in con­
nection with programming under uncertainty60.

Also, the method has been applied to PERT, where ran­
domness occurs in the times for the links associated with
the completion of a project, in connection with chance­
constrained programming61, and to economic development planning62.

CHAPTER VI

DUALITY IN STOCHASTIC PROGRAMS

INTRODUCTION

The theory of duality for stochastic linear programming is an extension of the duality theorem of the non-stochastic or deterministic programming problem. In this chapter the work done on the passive and active approaches with respect to duality will be presented. Research is still going on on the different aspects of duality in stochastic linear programming. The main effort is directed from a topological point of view in the parameter space of the stochastic program. The interesting aspect of duality here is that the regions of feasibility in the active and passive case may be different but the optimality regions of the primal and dual bases are identical.

I. THE PASSIVE CASE

Several results for the passive case presented in the previous chapter will be used here as a basis for proving theorems for the duals in the passive case.

Definition 6.1.1. If a particular region $S_k'$ and a selection $k'$ is found such that $Z^{k'} > Z_k$ in the whole region $S_k'$, and $k \neq k'$, then $S_k'$ is said to be the region of optimality for selection $k'$.
Definition 6.1.2 Under a given selection \( k \), from the results of the previous chapter, it can be concluded that the activity level vector \( X_i, (i=1,\ldots,n) \) will be undefined, where the determinant of the matrix formed by the selected columns takes on the value zero. The region \( V_k \) of the parameter space, where any variable \( X_i \) is undefined for the selection \( k \), is called a critical or singular region. For example, for \( k=1 \) from the example given in the previous chapter we have

\[
X_1^{(1)} = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \\ a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \frac{\Delta_1}{\Delta_2}
\]

If \( \Delta_2=0 \) then \( X_1^{(1)} \) belongs to the region \( V_1 \) which is singular.

Now, with these definitions out of the way, we can recall the duality theorem \(^{63}\) of the deterministic linear programming. It states that if a primal linear programming problem has a finite optimal solution, then so does the dual problem obtained from it, and furthermore, the maximum of the primal equals the minimum of the dual, in other words, their solutions are equal. Therefore we have:

Theorem 6.1.1 Let \( z^{(k)}(x) \) be the objective function determined by the \( k \)th selection in the solution of the primal stochastic linear program, and let \( U_k \) be the region in the

parameter space throughout which \( Z^{(k)}(x) \) is both optimal and feasible. Now, if \( q \) is any point in \( U(k) \), then there exists a \( k \)-th selection for the dual problem with corresponding objective function \( Z^{(n)}(x) \) such that \( Z^{(n)}(q) = Z^{(k)}(q) \) and a neighborhood \( N_{e}(q) \) about \( q \) throughout which \( Z^{(n)}(x) \) is both feasible and optimal, and such that \( Z^{(n)}(x) = Z^{(k)}(x) \) for all \( x \) in \( N_{e}(q) \).

**Proof:** According to the duality theorem, since \( q \) is a fixed point in the parameter space \( S \), and since the assumption is that the primal has a solution at \( q \), the dual must have a solution also. If this solution is given by the \( n \)-th selection, then \( Z^{(n)}(q) = Z^{(k)}(q) \). Now, there is a possibility that there may be more than one selection in the dual for which the above equation holds. Therefore, suppose that there is no neighborhood of \( q \) throughout which \( Z^{(n)}(x) \) is optimal, even though it is optimal at \( x=q \). Then since the \( k \)-th selection is feasible for the primal \( Z^{(k)}(x) \) is optimal and a solution to the primal throughout \( U_{k} \). Then suppose that the \( n \)-th selection for the dual is optimal at some points of an arbitrary neighborhood, say \( N_{e}(q) \) of \( q \), \( N_{e}(q) \subseteq U_{k} \). The result is that while

\[
Z^{(n)}(q) - Z^{(m)}(q) < 0,
\]

\[
Z^{(n)}(x) - Z^{(m)}(x) \not< 0 \text{ at some points in } N_{e}(q).
\]

If there are only a finite number \( x_1, x_2, \ldots, x_q \) of such points \( x \), then we may define \( f(x_i) = p(q, x_i) \), \( i = 1, \ldots, q \), where \( p \) is the metric on \( S \), and let \( e_i = \min f(x_i) \). Then \( N_{e}(q) \) is a neigh-

Now, let there be an infinite number of points $x_\lambda$, $\lambda$ in $\Delta$, where $\Delta$ is a countable or uncountable index set for which $Z^{(m)}(x_\lambda) \leq Z^{(n)}(x_\lambda)$. Now, since $$S = \bigcap_{i=1}^{\infty} n_i \cap \bigcup_{i=1}^{\infty} n_i$$
the $S$ is metric and first countable and there exists at $q$ a countable basis $N_{1/n}(q) \subset N_e(q)$ and in each $N_{1/n}(q)$ there exists at least one point $x_\lambda$.

The neighborhood $D_{1/n}(q)$ of $q$ can be defined now, as $D_{1/n}(q) = N_{1/n}(q) - N_{1/n+1}(q)$. Then there are two possibilities: either there are infinitely many $D_{1/n}(q)$, or there are only finitely many. If there are only finitely many, then there exists an integer $N > 0$ such that for all $N > N$, $D_{1/n}(q)$ does not contain any points $x_\lambda$, for all $n > N$ and $N_{1/n'}(q)$ forms a neighborhood of $q$ throughout which $Z^{(n)}(x) = Z^{(E)}(x)$, or there are infinitely many of the points $x_\lambda$ which belong to $N_{1/n}(q)$ for all $n > N$. These points for a set $TCS$, and $q$ is a limit point of $T$. Therefore, there is a sequence of distinct points of $T$ converging to $q$. Let this sequence be called $(x_{\lambda_n})$.

Since the $Z^{(m)}(x)$ are continuous at all points in $S$ except in a singular region, it follows that the sequence $Z^{(n)}(x_{\lambda_n})$ converges to $Z^{(n)}(q)$ for all $n$, and in particular $m$. But since $Z^{(m)}(q) - Z^{(n)}(q) \leq 0$ and $Z^{(m)}(x_{\lambda_n}) - Z^{(n)}(x_{\lambda_n}) \geq 0$ for all $n$, it follows that the sequence $W(x_{\lambda_n}) = Z^{(m)}(x_{\lambda_n}) - Z^{(n)}(x_{\lambda_n})$ does not converge to $W(q)$. This contradicts the continuity of $W(x)$. 


and therefore the continuity of \( \frac{1}{\gamma(n)}(X) \) and \( \frac{1}{\gamma(m)}(X) \).

If there are infinitely many \( D_{1/n}(q) \) containing points \( X_\lambda \), the axiom of choice can be used on each such \( D_{1/n}(q) \) to select out of each a point \( X_\lambda \), calling the point \( X_\lambda \) chosen out of \( D_{1/n}(q) \) as \( X_{1\lambda} \). The points \( (X_{1\lambda}) \) form a sequence which converges to \( q \). This proves the theorem.

II. THE ACTIVE CASE

Let \( U_{ij} \) be the decision variables in the active case of the linear stochastic model. Let the assumption be that the decision variables \( U_{11} \) and \( U_{22} \) can each assume at most a finite number of discrete values between 0 and 1 inclusive, e.g., let \( U_{11} \) assume \( P_1 \) values and \( U_{22} \) assume \( P_2 \) values. These values can be indexed by letting \( \psi = 1, \ldots, P_1 \) and \( \theta = 1, \ldots, P_2 \). If the problem has \( \lambda \) selections then let \( K = 1, \ldots, \lambda_1 \) be the index.

Definition 6.2.1. \( Z_{K}^{\psi, \theta}(x) \) is the objective function for the \( K \)th selection when \( U_{11} \) and \( U_{22} \) take on values given by \( \psi \) and \( \theta \) and \( X \) is any feasible point in the parameter space \( S \). Then the following theorem can be stated.

Theorem 6.2.1. Let \( \psi', \theta' \) and \( K' \) denote certain values of \( \psi, \theta \), and \( K \) respectively. Let \( Z_{K'}^{\psi', \theta'} \) be the corresponding objective function and let \( q \) be a point in the interior of a feasible region for the selection \( K' \) and for \( \psi' \) and \( \theta' \). Let us suppose that \( Z_{K'}^{\psi', \theta'}(q) > Z_{K}^{\psi, \theta}(q) \) for any combination of \( \psi, \theta \) and \( K \).
except $\varphi', \varepsilon'$, $K'$. Then there is a neighborhood $N_e(q)$ of $q$ such that for all $x \in N_e(q)$, $Z_{k'}^{\varepsilon'} (x) > Z_k^{\varepsilon} (x)$.

**Proof.** Suppose that there exists such a neighborhood of $q$, then, we have the fact that if $Z_{k'}^{\varepsilon'} (q) > Z_k^{\varepsilon} (q)$ then $Z_{k'}^{\varepsilon'} (x) > Z_k^{\varepsilon} (x)$ for infinitely many points $(x, \lambda)$ contained in an $e$-neighborhood $N_e(q)$, for every $e$. Let $N_{1/n} (q)$ be a spherical neighborhood of radius $1/n$ of $q$ and define $D_{1/n} (q) = N_{1/n} (q) - N_{1/n+1} (q)$. Now, if there are only finitely many of the $D_{1/n} (q)$ containing points $x, \lambda$ then the theorem is proved like the proof in theorem 6.1.1.

If there are infinitely many $D_{1/n} (q)$ containing points $x, \lambda$ then by the axiom of choice a point $x, \lambda$ is selected and the sequence $(x, \lambda_n)$ converges to $q$. The sequence $(H(x, \lambda_n)) = Z_{k'}^{\varepsilon'} (x, \lambda_n) - Z_k^{\varepsilon} (x, \lambda_n) does not converge to $H(q)$, since $H(q) > 0$, and $H(x, \lambda_n) < 0$ for all $n$. This implies that $H(x)$ is not a continuous function, this contradicts the assumed continuity of $Z_{k'}^{\varepsilon'} (x)$ and $Z_k^{\varepsilon} (x)$ outside the singular regions. This proves the theorem.

There is another interesting theorem in the active case by consideration of the dual. Let the objective function for the dual be given by $Z_h^{\varepsilon} (x)$ where $h=1, \ldots, \lambda$ if there are $\lambda$ selections to be made in the dual, then we have the following theorem.
Theorem 6.2.2. Let $Z_k^{\phi^\prime \theta^\prime}$ be the objective function for a certain solution and let $\phi$ and $\theta$ assume such values that, at some point $q$ in the feasible region of $k$th selection, $Z_k^{\phi^\prime \theta^\prime} (q) > Z_k^{\phi \theta} (q)$ for all $\phi, \theta$ and $k$. Then there is some neighborhood $N_{e}(q)$ about $q$ and a selection $h'$ for the dual and values $\phi''$ and $\theta''$ of $\phi$ and $\theta$ such that $Z_{h'}^{\phi'' \theta''} (x) = Z_k^{\phi^\prime \theta^\prime} (x)$ for all $x$ in $N_{e}(q)$.

Proof. According to the theorem of duality if $q$ is a fixed point in $S$, it is known that if $\phi'' = \phi^\prime$ and $\theta'' = \theta^\prime$ then there is a selection $h'$ for the dual such that $Z_{h'}^{\phi'' \theta''} (q) = Z_k^{\phi^\prime \theta^\prime} (q)$.

The values $\phi', \theta'$ and $h'$ suffice to make the theorem true at $x=q$.  

CHAPTER VII

CONCLUSIONS.

The problem of optimal decision-making under risk has been considered in this paper in the framework of the linear stochastic model, when the probability distributions of the objective function is sought.

The stochastic model described is constrained by three basic limitations:

(a) The assumption of normality of the distribution of random elements at one state or another of the analytical solution has the effect of making the results not distribution-free.

(b) The neglect of differing implications of alternative sample estimates of the parameters of $A$, $b$ and $c$ for the optimal solution vector $\bar{x}$.

(c) And the unexplored point of the interaction between prior and posterior probabilities.

These limitations are currently being investigated through extensive Monte Carlo computer simulations. These techniques utilize numerical procedures and they do not necessitate normality in the distribution of the random elements. If the random variables involved in the model are not independent, additional statistical problems arise and this aspect of the problem has not yet been investigated. A group of orthogonal transformations could be introduced into the model to make the random variables independent.

The problem of simultaneous variations in all the parameters of the model has not been solved yet. The present techniques accept only either variations in the vector
\( \mathbf{b}, \mathbf{c} \) (which is the same thing, since the presence of risk in the vector \( \mathbf{b} \) could very well be transferred to the coefficients of the objective function by solving the dual of the model.) or the matrix \( \mathbf{A} \). There is no practical theory for the simultaneous variations in all the parameters case. If the special cases of variations developed are generalized, the approximation formulas aggravate the problem of accuracy and the results become meaningless.

The present developments in the duality theory of the linear stochastic model have to be improved considerably if they are to be used in industrial planning applications.

The passive approach is very useful for development planning and analysis of the effect of random signal input into a system.

The active approach could be viewed as the decomposition of the restricting equations of the passive approach of stochastic programming, and this may prove to be very useful in optimal decision-making under decentralized organizational planning. If the data base of the model consists of time series, the empirical specification of the statistical distribution of the objective function, to different degrees of completeness and different choices of the allocation matrix \( \mathbf{U} \), may be useful for sequential aspects of macrodynamic policy-making and also for certain sequential test procedures.
In general, a great deal of research into the stochastic model is required yet to discover all the structural statistical properties of the distribution problem.
Selected Bibliography


Tintner, G. "A Note on Stochastic Linear Programming", *Econometrica*, XXVIII, 11(April 1960), 490-495


