

THE TRANSPORTATION PROBLEM

by

DALE GARRATT SHAW

B. A., Colorado College, 1963

445

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:

S. Thomas Parker
Major Professor

LD
2668
R4
S 534
C. 2

TABLE OF CONTENTS

INTRODUCTION	1
FORMULATION OF THE PROBLEM	1
DEFINITIONS AND THEOREMS	9
INITIAL SOLUTIONS	22
THE DISTRIBUTION METHODS	38
DEGENERACY	52
UNEQUAL SUPPLY AND DEMAND	57
OTHER RESTRICTIONS	59
ALTERNATE SOLUTIONS	60
CONCLUSION	66
ACKNOWLEDGMENT	67
BIBLIOGRAPHY	68

INTRODUCTION

A particularly interesting special case of the general linear programming problem is the transportation problem. This special problem is important from a practical as well as a theoretical standpoint. A sizable fraction of the applications of existing linear programming methods has been made in solving transportation and related problems. The theoretical importance of the transportation problem stems from the fact that many of the computational procedures which have been developed for its solution are, for the most part, simplifications of classical linear programming methods.

This paper contains a detailed discussion of so-called distributive methods for obtaining a solution for any given transportation problem. These methods were originally derived from the mathematical theory of the simplex method of linear programming. The transportation theory underlying these methods, however, can be developed independently of the programming theory of the simplex method. The intent of this paper is to construct this independent development, and to give some indication as to the efficacy of the various methods.

FORMULATION OF THE PROBLEM

A statement of the transportation problem can be made as follows: Determine a shipping schedule to "transport" a homogeneous product from various "origins" to various "destinations"

at a minimum total cost. The supply at each origin, the total requirements of each destination, and the costs to ship goods from each origin to each destination are known.

To construct a model to solve the transportation problem mathematically, it is necessary to make several limiting assumptions. Since it is desirable in all methods for obtaining a solution, to have the costs given in terms of so much per unit, the "proportionality" assumption is made. It is assumed that if a cost of c dollars is involved in shipping one unit from origin i to destination j , the cost to ship k units along the same route will be exactly k times as much, or kc dollars. This assumption leads to the assumption of the homogeneity of the product in the statement of the problem. The proportionality assumption requires the units of all the products to be shipped to be the same. In other words, the product is to be homogeneous.

An analysis of the problem leads to three more restrictions to be made, all of which may be stated as assumptions necessary to construct the model. If the flow of goods is permitted in only one direction, it is physically impossible to ship a negative quantity of units. This is the "nonnegativity" assumption, requiring the number of units shipped from origin i to destination j to be nonnegative. For the practical purpose of accounting, it is necessary to assume that any shipment of the product will neither create nor destroy it. Thus the total supply of the product distributed among the origins before any of the product is shipped will equal the sum of the amounts at the destinations after shipment. This is known as the "additivity"

assumption.

Finally it is noted that the total cost of a shipping schedule is a linear function of the number of units to be shipped from each origin to each destination. Multiplying the number of units to be shipped from each origin i to each destination j by the cost per unit to ship by that route, and summing these costs for all routes used yields the total cost of the shipping schedule. Conceivably a different method for determining the total cost could be found which would not be a linear function of the number of units shipped from each origin to each destination. In this paper the "linearity" of the total cost function, or "object" function, as it is sometimes called, is assumed.

A mathematical statement of the problem can now be made based on the above assumptions. Notation and terminology are necessarily introduced at this point.

Let O_i denote the i^{th} origin
 D_j denote the j^{th} destination
 m be the number of origins
 n be the number of destinations
 a_i be the number of units to be shipped from O_i
 b_j be the number of units required to be shipped
to D_j
 x_{ij} be the number of units to be shipped from O_i
to D_j
 c_{ij} be the cost to ship one unit from O_i to D_j
 X be a solution matrix composed of x_{ij} for all i, j

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1n} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_{m1} & x_{m2} & \cdot & \cdot & \cdot & x_{mn} \end{bmatrix}$$

Z be the total cost associated with the solution matrix X

A be the total volume of goods to be shipped.

The matrix X represents a possible shipping schedule. If X satisfies all of the conditions imposed by the assumptions it will be a solution in the sense that it describes a shipping program.

Looking first at the possible values each x_{ij} may assume, denote any limitation which can be placed on them as a constraint. The first constraint

$$x_{ij} \geq 0 \quad \text{for all } i, j$$

is taken from the nonnegativity assumption. Since there are a_i units to be shipped from O_i , each solution matrix X must satisfy

$$\sum_{j=1}^n x_{ij} = a_i \quad \text{for all } i$$

and since each D_j must receive b_j units, X must also satisfy

$$\sum_{i=1}^m x_{ij} = b_j \quad \text{for all } j$$

Any matrix X whose elements satisfy these constraints is a solution. Without further specifications, there may be many solutions.

The cost associated with each matrix X is given by

$$\begin{aligned}
 Z = & c_{11}x_{11} + c_{12}x_{12} + \dots + c_{1n}x_{1n} \\
 & + c_{21}x_{21} + c_{22}x_{22} + \dots + c_{2n}x_{2n} \\
 & \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\
 & \quad \cdot \quad \quad \cdot \quad \quad \quad \cdot \\
 & + c_{m1}x_{m1} + c_{m2}x_{m2} + \dots + c_{mn}x_{mn}
 \end{aligned}$$

or, in more compact form,

$$Z = \sum_{i=1}^m \sum_{j=1}^n c_{ij}x_{ij}$$

Since minimum total cost is the objective in the problem, it is necessary to find a smallest Z, which can be stated

$$\text{minimize } Z = \sum_{i=1}^m \sum_{j=1}^n c_{ij}x_{ij}$$

At no time has it been specified that the supply be exactly equal to the demand, although it was implied in making the additivity assumption. It is now convenient to make this stipulation in that it is necessary to lay the theoretical groundwork for the various computational methods which follow. The equality of supply and demand is not stated as a general assumption, however, since each method can readily accept a situation in which they are unequal. The following additional condition can thus be introduced:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j = A .$$

This condition can be thought of as a consistency condition which must be satisfied if a solution is to exist.

The mathematical statement of the problem is made as follows.

$$\text{Determine } X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1n} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_{m1} & x_{m2} & \cdot & \cdot & \cdot & x_{mn} \end{bmatrix}$$

such that

$$x_{ij} \geq 0 \quad \text{for all } i, j \quad (1)$$

$$\sum_{j=1}^n x_{ij} = a_i \quad \text{for all } i \quad (2)$$

$$\sum_{i=1}^m x_{ij} = b_j \quad \text{for all } j \quad (3)$$

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} = Z \text{ is a minimum}$$

$$\text{and} \quad \sum_{i=1}^m a_i = \sum_{j=1}^n b_j = A. \quad (4)$$

Writing out (2) and (3),

$$\begin{array}{rcl}
 x_{11}+x_{12}+\dots+x_{1n} & & = a_1 \\
 & x_{21}+x_{22}+\dots+x_{2n} & = a_2 \\
 & \cdot & \cdot \\
 & \cdot & \cdot \\
 & \cdot & \cdot \\
 & & x_{m1}+x_{m2}+\dots+x_{mn} = a_m \quad (5) \\
 x_{11} & & +x_{21} & \dots & x_{m1} & = b_1 \\
 & x_{12} & & +x_{22} & \dots & +x_{m2} & = b_2 \\
 & \cdot & & \cdot & & \cdot & \cdot \\
 & \cdot & & \cdot & & \cdot & \cdot \\
 & \cdot & & \cdot & & \cdot & \cdot \\
 & & x_{1n} & & +x_{2n} \dots & & +x_{mn} = b_n
 \end{array}$$

one observes that this is a system of $m+n$ equations in mn unknowns. On writing (5) in matrix form, and using (4), it is seen that the sum of the first m rows is equal to the sum of the last n rows. Therefore any row can be considered a linear combination of the remaining $m+n-1$ rows. Thus the system reduces to $m+n-1$ equations in mn unknowns. Although this fact is necessary in the next section, it is convenient to retain all the equations for computational purposes. The methods to be presented are best worked using a chart form as in Fig. 1 to represent all the necessary data. All the information necessary to work a transportation problem having three origins and three destinations appears here. The left column identifies each origin and the top row identifies each destination. The c_{ij} in the upper right corner of each interior square common to O_i and D_j is the cost coefficient corresponding to that route. Each square, then, can be thought of as the route connecting some O_i and some D_j along which an assignment of x_{ij} units is shipped.

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
O_1		c_{11}		c_{12}		c_{13}		a_1
	x_{11}		x_{12}		x_{13}			
O_2		c_{21}		c_{22}		c_{23}		a_2
	x_{21}		x_{22}		x_{23}			
O_3		c_{31}		c_{32}		c_{33}		a_3
	x_{31}		x_{32}		x_{33}			
Req.		b_1		b_2		b_3		A

Fig. 1.

The last row and column indicate the supplies available at each origin and the requirements of each destination. These entries will occasionally be referred to as the "rim" conditions. The figure in the lower right corner is the total volume of goods to be shipped and will equal the sum of the origin capacities as well as the sum of the destination requirements.

While it is true that transportation theory is an important part of the more general linear programming theory and can be developed from this theory, it is not the purpose of this paper to show this development. It suffices to indicate that the transportation problem is a linear programming problem and can be worked by the simplex method. Stated in simplex form the problem would be: Determine

Definition 1. A solution to the transportation problem is a matrix X

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

which satisfies the constraints

$$\sum_{i=1}^m x_{ij} = b_j \quad \text{for } j = 1, 2, \dots, n$$

$$\sum_{j=1}^n x_{ij} = a_i \quad \text{for } i = 1, 2, \dots, m$$

The physical interpretation of the variable x_{ij} has been said to be simply a shipment of a number of goods from O_i to D_j . It may also be called an allocation or assignment of some or all of the goods at O_i to D_j .

Definition 2. A solution variable is a variable which is nonzero in a solution.

Definition 3. A feasible solution is a solution in which all of the solution variables are positive.

Definition 4. A basic solution is a solution obtained by setting $mn - m - n + 1$ of the variables equal to zero and solving for the remaining $m + n - 1$ variables.

Definition 5. A basis is the collection of the $m + n - 1$ variables which are not set equal to zero in the construction of a basic solution.

Definition 6. A basic feasible solution is a feasible solution with no more than $m+n-1$ of the solution variables being positive.

Definition 7. A minimum feasible solution is a solution which satisfies

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij}x_{ij} = Z \text{ is a minimum.}$$

Definition 8. A nondegenerate basic feasible solution is a basic feasible solution with exactly $m+n-1$ solution variables being positive. A degenerate basic feasible solution has fewer than $m+n-1$ solution variables.

The following examples illustrate the connections between the above definitions and the information as can be represented in the chart form. Consider the solution as represented in chart form in Fig. 2.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		20		7		4		60
	50		10					
O ₂		17		3		6		40
			40					
O ₃		15		10		6		60
			20		40			
Req.		50		70		40		160

Fig. 2.

The solution matrix X is

$$X = \begin{bmatrix} 50 & 10 & 0 \\ 0 & 40 & 0 \\ 0 & 20 & 40 \end{bmatrix}$$

This X represents a solution which is interpreted as

$$\begin{array}{lll} x_{11} = 50 & x_{12} = 10 & x_{13} = 0 \\ x_{21} = 0 & x_{22} = 40 & x_{23} = 0 \\ x_{31} = 0 & x_{32} = 20 & x_{33} = 40 \end{array}$$

indicating this solution or program calls for a shipment of 50 units from O_1 to D_1 , a shipment of 10 units from O_1 to D_2 , and so on. The cost of this program is given by

$$\begin{aligned} Z &= \sum_{i=1}^3 \sum_{j=1}^3 c_{ij}x_{ij} \\ &= c_{11}x_{11} + c_{12}x_{12} + c_{13}x_{13} + c_{21}x_{21} + c_{22}x_{22} \\ &\quad + c_{23}x_{23} + c_{31}x_{31} + c_{32}x_{32} + c_{33}x_{33} \\ &= 20 \times 50 + 7 \times 10 + 4 \times 0 + 17 \times 0 + 3 \times 40 \\ &\quad + 6 \times 0 + 15 \times 0 + 10 \times 20 + 6 \times 40 \\ &= 1630 \end{aligned}$$

As seen by an inspection of the chart, the constraints

$$\sum_{i=1}^3 x_{ij} = a_i \quad \text{for all } j = 1, 2, 3$$

$$\sum_{j=1}^3 x_{ij} = b_j \quad \text{for all } i = 1, 2, 3$$

are satisfied and thus X is a solution.

The solution variables are x_{11} , x_{12} , x_{22} , x_{32} , and x_{33} .

The solution is a feasible one since these are all positive. The solution is also a basic feasible solution since $m+n-1 = 3+3-1 = 5$ and there are no more than 5 solution variables. Since there are exactly 5 solution variables, the solution is also a nondegenerate one.

The chart in Fig. 3 indicates a degenerate basic feasible solution for the same problem. Notice that only 4 solution

	⋮	D_1	⋮	D_2	⋮	D_3	⋮	Supply
O_1		20		7		4		60
	50		10					
O_2		17		3		6		40
					40			
O_3		15		10		6		60
			60					
Req.	50		70			40		160

Fig. 3.

variables appear here, thus making the solution degenerate. This degenerate solution was possible because a partial sum of the row requirements was equal to a partial sum of the column requirements. In this case $a_2 = b_3$. If more than one set of partial sums are equal, multiple degeneracy can occur. In this case a solution with even fewer solution variables would exist.

It is desirable, as will be seen in the next section, to work with problems in which only nondegenerate solutions can exist. Theorem 1 indicates one method of avoiding the possibility of obtaining a degenerate solution at any stage in the algorithms to follow.

Theorem 1. For every problem in which degenerate solutions can occur, there are equivalent problems in which degenerate basic feasible solutions are impossible.

Proof. Degeneracy can occur only when partial sums of row and column requirements are equal. An adjustment which makes this partial equality impossible also avoids any possibility of degeneracy. The method which will be used throughout this paper is to add ϵ to each column requirement and add $n\epsilon$ to the last row requirement, where ϵ is a positive number but smaller than any solution variable could ever be. This ϵ is called a perturbation constant. The condition

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

is still satisfied since the addition of $n\epsilon$ is made to both sides of the equation. Now a partial sum over r of the column requirements will contain the infinitesimal part $r\epsilon$, and any partial sum of the row requirements, 0ϵ or $n\epsilon$. The only case in which row requirements and column requirements will be equal then, occurs when all of the rows and columns are included. Thus a partial sum of the row requirements can never be equal to a partial sum of the column requirements, and hence degeneracy cannot occur. The original problem can be re-established at

any time by allowing ϵ to approach zero.

From this point on it will be assumed that there is no possibility of obtaining a degenerate basic feasible solution. This allows much simplification in the proofs of the following theorems.

Theorem 2. A basic feasible solution for any transportation problem always exists. In other words, a solution with at most $m + n - 1$ positive solution variables can be found for every problem.

Proof. Such a solution can be obtained step by step as follows.

1. Set any variable $x_{rs} = \text{Min}(a_r, b_s)$.
2. If $a_r > b_s$, then $x_{rs} = b_s$, and all other x_{ij} in this column must necessarily be zero to satisfy the column requirement. Delete column s from the matrix and continue in the same manner with the reduced matrix.
3. If $b_s > a_r$, then $x_{rs} = a_r$, and all other x_{ij} in this row are set equal to zero. Delete row r from the matrix and continue in the same manner with the reduced matrix.
4. This process is continued until the solution is complete.

The case where $a_r = b_s$ is not considered. It will only occur when the last assignment is made, since nondegeneracy of a solution has been assumed. The determination of each positive x_{ij} eliminates either a row or a column one at a time with the last assignment eliminating both a row and a column. It follows,

then, that this solution will have exactly $m + n - 1$ solution variables. Each rim condition is satisfied and thus the solution is a nondegenerate basic feasible solution.

Theorem 3. Assuming that a basic feasible solution has been found, a second basic feasible solution can be constructed from the first by introducing a new variable into the solution and removing a solution variable from it.

Proof. Suppose an initial basic feasible solution has been obtained and represented in chart form as in Fig. 4.

	⋮	D_1	⋮	D_2	⋮	D_3	⋮	Supply
O_1		20		7		4		60
	50		10					
O_2		17		3		6		40
			40					
O_3		15		10		6		60
			20		40			
Req.	50		70		40			160

Fig. 4.

Suppose also that x_{21} will be the new variable to be put into the program at some positive but small value θ . If θ is added to x_{21} , it must also be subtracted from x_{11} and x_{22} in order to continue to satisfy the rim conditions b_1 and a_2 . Notice now that the addition of θ to x_{12} will simultaneously satisfy the

rim conditions b_2 and a_1 which were left unsatisfied by the subtraction of θ from x_{11} and x_{22} , and that all rim conditions are again satisfied. Figure 5 shows the new program with the introduction of θ in x_{21} . The sequence of variables traced by the

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
0_1		20		7		4		60
		$50 - \theta$		$10 + \theta$				
0_2		17		3		6		40
		θ		$40 - \theta$				
0_3		15		10		6		60
				20		40		
Req.		50		70		40		160

Fig. 5.

additions and subtractions of θ is called a path.

If θ is now increased from an initial value close to zero, the solution variables from which θ is being subtracted become smaller as θ becomes larger. If θ is allowed to increase until one of these variables is reduced to zero, this new solution will contain $m + n - 1$ solution variables. If θ became any larger, this smallest solution variable would become negative and the solution would no longer be a basic feasible one. This new basic solution must always exist since any introduction of

θ to a nonbasic variable necessitates its subtraction from at least the two solution variables in the same row and column. Thus at least two solution variables will be driven toward zero. The assumption of nondegeneracy is assurance that no two of these decreasing variables go to zero at the same time.

In addition this new solution is unique. The basis variables will all reach zero at some time or other with the appropriate increase in θ . However, as soon as the first reaches zero, θ can get no larger since this basis variable would become negative. Thus the introduction of θ in a nonbasic variable will drive out one and only one unique solution variable. Hence the new solution is a unique derivation from the starting solution.

At this point, an initial basic feasible solution for every transportation problem is guaranteed to exist. Also a new basic feasible solution can be constructed from the first one. Indications later will be given as to what variables should be introduced so as to give a better solution in the sense that the new solution will be of lower cost.

The last theorem is by far the most important in transportation theory. It gives a reason for dealing only with basic feasible solutions in the previous theorems. The statement of this theorem requires that the problem have a unique optimal solution, although this is often not the case. A discussion of this appears in the section dealing with alternate solutions.

Theorem 4. Given a transportation problem which has a unique optimum, the minimal solution must be a basic feasible solution.

Proof. Assume all possible basic solutions have been found and of these the solution in Fig. 6 is found to be of minimal cost.

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
C_1		20		7		4	60
			30		30		
O_2		17		3		6	40
			40				
O_3		15		10		6	60
	50				10		
Req.		50		70		40	160

Fig. 6.

A similar solution in Fig. 7 is one in which x_{11} has been increased from zero to θ , a small positive number, to give $m + n$ solution variables.

The first of these solutions is seen to be a basic feasible solution since it does contain five solution variables and all of the rim requirements are satisfied. Let this solution be X_1 with an associated cost of Z_1 . The second solution is not a basic solution since it has six solution variables. It will be denoted by X_2 with a cost of Z_2 .

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
0_1	θ	20		7		4		60
			30		30			
0_2		17		3		6		40
			40					
0_3	50	15		10		6		60
					10			
Req.		50		70		40		160

Fig. 7.

The assumption that Z_2 is less than Z_1 is made and shown to lead to a contradiction. If some X_2 could be found such that Z_2 is less than Z_1 , this would indicate that some nonbasic solution could be better than any basic solution; however, this will be shown not to be the case.

In a manner similar to the method used in the example in Theorem 3, the path for θ in the new solution is found. The introduction of θ in x_{11} necessitates its subtraction from x_{13} and x_{31} and its addition to x_{33} to keep the rim requirements satisfied. This second program is shown in Fig. 8. If θ is increased from a small positive quantity, approximately zero, to the value one, the expression $c_{11} - c_{13} + c_{33} - c_{31}$ will represent the change in the cost of the solution for the introduction of one unit of x_{11} . This is called the opportunity cost

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
·		20		7		4		60
O ₁	θ		30			30 - θ		
		17		3		6		40
O ₂			40					
		15		10		6		60
O ₃	50 - θ					10 + θ		
Req.		50.		70		40		160

Fig. 8.

of x_{11} and will be denoted by O_{11} . Thus

$$O_{11} = c_{11} - c_{13} + c_{33} - c_{31}.$$

If $Z_2 < Z_1$, then O_{11} must necessarily be negative, with the cost of X_2 being given by

$$Z_2 = Z_1 + \theta O_{11} \quad (6)$$

Now suppose θ is allowed to increase until one of the old solution variables becomes zero. This new solution, X_3 , will also be a basic feasible solution. If the smallest decreasing variable becomes zero at $\theta = \theta'$, the cost of this third solution is

$$Z_3 = Z_1 + \theta' O_{11}$$

and $Z_3 < Z_1$ since (6) implies that the solution cost decreases as θ increases. However, X_3 is a basic feasible solution, and therefore its cost Z_3 must be greater than Z_1 . Hence a contradiction has been obtained, and therefore the assumption $Z_2 < Z_1$

must be false.

Beginning with X_1 again and introducing two extra nonzero variables into the solution matrix, the assumption that the cost of this new solution, X_4 , is less than that of X_1 is made. If either of the extra variables is increased until an old solution variable becomes zero, this new solution, X_5 , will have $m + n$ solution variables. However, such a solution was just shown not to be minimal and so the assumption that $Z_4 < Z_1$ is false. This argument can be extended to any number of extra solution variables, thus proving the theorem.

Theorem 4, then, is assurance that no nonbasic solution can exist which has a cost less than that of the optimal basic solution.

Although the proofs of Theorems 3 and 4 are based on a sample problem having three origins and three destinations, this in no way affects their generality. Neither proof is dependent upon the size of the problem nor upon the choice of the extra variables introduced into a solution.

INITIAL SOLUTIONS

The first step in each of the following methods for solving a transportation problem is to obtain an initial program from which to start. A solution matrix X must be obtained such that it is a basic feasible solution, or, in other words, a program is to be found in which $m + n - 1$ of the variables are positive. There are various methods available to obtain this initial

program of which five are mentioned here. The three most commonly used are outlined in detail and an example for each given.

The Northwest Corner Rule for obtaining an initial program is probably the most computer orientated. No time is spent searching for the various elements of the matrix since no attention is paid to the cost coefficients associated with each route. This method does not usually give as good an initial solution as can be obtained by other methods.

The scheme is as follows.

1. Let $x_{11} = \text{Min}(a_1, b_1)$.
2. Reduce both a_1 and b_1 by x_{11} .
3. Repeat steps 1 and 2 using the upper left-hand (north-west) corner element of the matrix obtained by deleting the row or column whose rim requirement has just been satisfied.
4. This process is repeated until the solution is complete.

The assignments are made by starting in the upper left-hand corner of the given transportation matrix and exhausting each origin capacity and satisfying each destination requirement one at a time until the last assignment simultaneously satisfies the requirements of the last row and column. There will be exactly $m + n - 1$ assignments made since there are n rows, m columns, and the last assignment simultaneously satisfies both a row and column requirement. Thus the Northwest Corner Rule yields a basic feasible solution. It happens, however, that some of the assignments, other than the last one, satisfy the requirements of some destination and exhaust the capacities of some origin at

the same time. This indicates that degeneracy has occurred and the solution obtained is degenerate. If this happens, the problem should be restated with the addition of a perturbation constant to avoid any further possibility of obtaining a degenerate solution. It is usually best to add the perturbation constant before the problem is begun, rather than wait until degeneracy is encountered during its solution. The time saved in not having to begin over again is often worth the extra computations involved in carrying along the perturbation constant.

The following example is introduced here to facilitate the understanding of the Northwest Corner Rule. It will also be used throughout the rest of this paper in an attempt to evaluate somewhat the efficacy of the various methods.

Example 3. Table 1 gives all of the information for solving a transportation problem with 3 origins and 3 destinations.

Table 1.

Origin capacities		Destination requirements		Cost coefficients			
					D_1	D_2	D_3
O_1	60 units	D_1	50 units	O_1	20	7	4
O_2	40 units	D_2	80 units	O_2	17	3	6
O_3	60 units	D_3	30 units	O_3	15	10	6

The cost to ship one unit from O_i to D_j is read from the square common to the i^{th} row and j^{th} column in the Cost Coefficients Table. Figure 9 is the representation of the above data in chart form. Following the steps of the Northwest Corner Rule,

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	60
O_2		17		3		6	40
O_3		15		10		6	60
Req.		50		80		30	160

Fig. 9.

x_{11} is set equal to

$$\text{Min}(a_1, b_1) = \text{min}(60, 50) = 50$$

and is subtracted from both a_1 and b_1 , leaving $a_1 = 10$ and $b_1 = 0$. See Fig. 10. Since $b_1 = 0$, the first column is ignored

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1	50	20		7		4	10
O_2		17		3		6	40
O_3		15		10		6	60
Req.		0		80		30	

Fig. 10.

and the new northwest corner element is x_{12} . Now $x_{12} = \text{Min}(a_1, b_2) = \text{min}(10, 80) = 10$ and is subtracted from both a_1 and b_2 , leaving $a_1 = 0$ and $b_2 = 70$. Rather than physically deleting the first column from the matrix, it is more convenient to leave it in, noting that the new entry of $b_1 = 0$ can signify that this column will no longer be used. The chart in Fig. 11 results after making the second assignment of $x_{12} = 10$. The new

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	0
	50		10				
O_2		17		3		6	40
O_3		15		10		6	60
Req.	0		70			30	

Fig. 11.

northwest corner element now is x_{22} , thus $x_{22} = \text{Min}(40, 70) = 40$ and $a_2 = 0$ and $b_2 = 30$. Following this assignment, x_{32} becomes the new northwest corner element, and hence $x_{32} = \text{Min}(60, 30) = 30$ and $a_3 = 30$ and $b_2 = 0$. These last two assignments are indicated in the chart in Fig. 12. At this point, 4 assignments have been made and the only square left for consideration is the one common to the third row and third column.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
0 ₁		20		7		4		0
	50		10					
0 ₂		17		3		6		0
			40					
0 ₃		15		10		6		30
			30					
Req.		0		0		30		

Fig. 12.

It becomes the new northwest corner element and $x_{33} = \text{Min}(a_3, b_3) = \text{min}(30, 30)$. Note that both a_3 and b_3 are satisfied by this last assignment. There have been 5 assignments made satisfying all of the rim requirements, and thus this is a basic feasible solution.

Filling in the original chart with all of the assignments (see Fig. 13), indicates the complete solution is

$$\begin{array}{lll}
 x_{11} = 50 & x_{12} = 10 & x_{13} = 0 \\
 x_{21} = 0 & x_{22} = 40 & x_{23} = 0 \\
 x_{31} = 0 & x_{32} = 30 & x_{33} = 30
 \end{array}$$

The cost associated with this solution will be

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
0 ₁		20		7		4		60
	50		10					
0 ₂		17		3		6		40
			40					
0 ₃		15		10		6		60
			30		30			
Req.		50		80		30		160

Fig. 13.

$$\begin{aligned}
 Z &= \sum_{i=1}^3 \sum_{j=1}^3 c_{ij}x_{ij} \\
 &= c_{11}x_{11} + c_{12}x_{12} + c_{13}x_{13} + c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23} \\
 &\quad + c_{31}x_{31} + c_{32}x_{32} + c_{33}x_{33} \\
 &= 20 \times 50 + 7 \times 10 + 4 \times 0 + 17 \times 0 + 3 \times 40 + 6 \times 10 \\
 &\quad + 15 \times 0 + 10 \times 30 + 6 \times 30 \\
 &= 1810
 \end{aligned}$$

A restatement of the problem with the addition of a perturbation constant was unnecessary for this example since all of the assignments made were positive.

It is interesting to note that had D₁ and D₃ been interchanged, the Northwest Corner Rule would have given the solution in Fig. 14. For this solution, Z = 1300, a considerable savings over the cost of the first solution. This solution is only one

	\vdots	D_3	\vdots	D_2	\vdots	D_1	\vdots	Supply
O_1		20		7		20		60
	30		30					
O_2		17		3		17		40
			40					
O_3		15		10		15		60
			10		50			
Req.		30		80		50		160

Fig. 14.

iteration away from the optimal solution, whereas the first initial solution will require 3 iterations before the optimal solution is obtained. It is readily apparent now that the Northwest Corner Rule is essentially the same method as was constructed in Theorem 2, with the added restriction of making the first assignment in the northwest corner square. There is no guarantee that the solutions obtained by either of these methods will be anywhere near the optimal solution. Their use could conceivably give the "worst" possible solution, in the sense of being the solution of highest cost, or even the optimal solution with no further iterations necessary.

The next two methods take into consideration the cost coefficients and tend to make assignments along routes involving the lower given costs. The first is appropriately called the

Least Cost Rule. The cost coefficient matrix is scanned for the smallest entry. The x_{ij} corresponding to the smallest cost coefficient is used for the first assignment.

The steps can be stated as follows.

1. Scan the cost matrix for the smallest c_{ij} .
2. Set $x_{ij} = \text{Min}(a_i, b_j)$ and reduce a_i and b_j both by x_{ij} .
3. Repeat step 2 using the smallest element in the cost matrix obtained by deleting the row or column already satisfied.
4. This process is repeated until the solution is complete.
5. In the case of a tie for the smallest c_{ij} any arbitrary rule may be used to break the tie.

The same problem of degeneracy occurs here also if any assignment other than the last one satisfies both a row and column requirement at the same time.

To illustrate this method with Example 3 used before, the cost coefficients are first scanned and $c_{22} = 3$ is found to be the smallest. So $x_{22} = \text{Min}(a_2, b_2) = \text{min}(40, 80) = 40$ and $a_2 = 0$ and $b_2 = 40$. The chart with this first assignment appears in Fig. 15. Row 2 is now ignored since its requirements are met and the reduced cost matrix is scanned for smallest cost coefficient which is $c_{13} = 4$. Therefore $x_{13} = \text{Min}(30, 60) = 30$ and $b_1 = 30$ and $a_1 = 0$. See Fig. 16. The requirement in column 3 is now satisfied, the column ignored, and scanning the cost matrix finds the smallest c to be $c_{12} = 7$. So $x_{12} = \text{Min}(40, 30) = 30$ and $a_1 = 0$ and $b_2 = 10$. Row 1 is now ignored and $c_{32} = 10$ is the smaller of the two remaining cost coefficients. Thus

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	60
O_2		17		3		6	0
			40				
O_3		15		10		6	60
Req.		50		40		30	

Fig. 15.

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	30
					30		
O_2		17		3		6	0
			40				
O_3		15		10		6	60
Req.		50		40		0	

Fig. 16.

$x_{32} = \text{Min}(60, 10) = 10$ and $a_3 = 50$ and $b_1 = 0$. The last assignment is $x_{31} = 50$ and the complete solution in chart form appears

in Fig. 17. The cost of this solution is

$$Z = 7 \times 30 + 4 \times 30 + 3 \times 40 + 15 \times 50 + 10 \times 10 = 1300.$$

Five assignments were again made, so this is a basic feasible solution.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		20		7		4		60
			30		30			
O ₂		17		3		6		40
			40					
O ₃		15		10		6		60
	50		10					
Req.		50		80		30		160

Fig. 17.

It is interesting to note here that a rearrangement of rows or columns will not affect the cost obtained by the Least Cost Rule, whereas a change did affect the Northwest Corner Rule.

None of the initial solutions thus far obtained was optimal. The most desirable method, of course, would guarantee an optimal solution on the first try; however, no such method is in existence yet today. Vogel's Approximation Method (VAM) to be considered next is considered better than the two initial approximation methods previously presented in the sense that its

initial solution is usually "reasonably" close to the optimal one.

In Vogel's Approximation Method, a difference column and a difference row, representing the differences between the cost coefficients of the two cheapest routes for each origin and destination are computed. Each individual difference can be thought of as a penalty for not using the cheaper of the two routes. The highest penalty rating is identified and the first assignment is made to the x_{ij} corresponding to the smallest cost coefficient in that row or column.

The procedure can be stated as follows.

1. Compute the penalty ratings for each row and column, select the largest, and identify its row or column.
2. The x_{ij} which corresponds to the smallest cost in that row or column is set equal to $\text{Min}(a_i, b_j)$ and is subtracted from a_i and b_j .
3. Recompute the penalty ratings, again choose the largest, and repeat step 2 using the matrix obtained by deleting the first row or column already satisfied.
4. This process is repeated until the solution is complete.
5. In the case of a tie for the largest penalty coefficient choose the smallest c_{ij} in the tied rows or columns.

In looking at Example 3 again one more row and another column are added to the chart to facilitate the computation of the penalty ratings. The differences between the smallest two

cost coefficients in each row and column are recorded in the ratings row and column. See Fig. 18.

	D_1	D_2	D_3	Supply	Row ratings
O_1	20	7	4	60	3
O_2	17	3	6	40	3
O_3	15	10	6	60	4
Req.	50	80	30	160	
Column ratings	2	4	2		

Fig. 18.

In this case there is a tie for the largest penalty rating. Both ratings in row 3 and column 2 are 4. The smallest c in either row 3 or column 2 is $c_{22} = 3$. Therefore $x_{22} = \text{Min}(a_2, b_2) = \text{min}(40, 80) = 40$, a_2 and b_2 are each reduced by 40, the second row is ignored, and new penalty ratings are computed. See Fig. 19. The largest penalty coefficient is 5 in column 1, $c_{31} = 15$ is the smallest cost coefficient in that column, and therefore $x_{31} = \text{Min}(60, 50) = 50$. Now $a_3 = 10$, $b_1 = 0$, column 1 is ignored, and new penalty ratings are computed. See Fig. 20. The largest penalty rating is 4 in row 3 and $c_{33} = 6$ is the smallest cost coefficient in that row. Thus

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply	⋮	Row ratings
O ₁		20		7		4		60		3
O ₂		17		40	3		6	0		-
O ₃		15		10		6		60		4
Req.	50		40		30					
Column ratings	5		3		2					

Fig. 19.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply	⋮	Row ratings
O ₁		20		7		4		60		3
O ₂		17		40	3		6	0		-
O ₃	50	15		10		6		10		4
Req.	0		40		30					
Column ratings	-		3		2					

Fig. 20.

$x_{33} = \text{Min}(10, 30) = 10$ and $a_3 = 0$ and $b_3 = 20$. The chart with the addition of this last assignment appears in Fig. 21.

	∴	D ₁	∴	D ₂	∴	D ₃	∴	Supply	∴	Row ratings	
O ₁		20		7		4		60		3	
O ₂		17		40		3		6		0	-
O ₃	50	15		10		10		6		0	-
Req.	0			40		20					
Column ratings	-			-		-					

Fig. 21.

At this point the column penalty ratings cannot be computed as only one cost coefficient remains in each column. However, there is only one way to finish making the assignments and still satisfy all of the rim requirements. The remaining assignments, $x_{12} = 40$ and $x_{13} = 20$, are made and the complete solution appears in Fig. 22. This solution has a cost of 1290.

It is illustrative at this point to compare the costs of the solutions obtained by the three methods for Example 3. For the solution obtained by the Northwest Corner Rule, $Z = 1810$ was the associated cost. A rearrangement of the given data produced a cost of 1300. The Least Cost Rule and VAM arrived at

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	60
			40		20		
O_2		17		3		6	40
			40				
O_3		15		10		6	60
	50				10		

Fig. 22.

solutions with costs of 1300 and 1290 respectively. It will soon be shown that the VAM solution is actually the optimal solution for this example, indicating that possibly this method is superior to the other two.

There are other methods for obtaining an initial basic feasible solution. The Row Minimum method assigns as much as possible to the x_{ij} corresponding to the smallest cost coefficient in the first row consistent with the rim requirements. The appropriate row or column is deleted and the process repeated using the first row of the new matrix until the solution is completed. The Column Method is similar. Other methods seem to be little more than inspection methods with varying degrees of organization.

THE DISTRIBUTION METHODS

Two iterative methods for obtaining an optimal solution to any transportation problem are presented in this section. Not all methods are iterative, as was indicated in Theorem 4. Any possible solutions to Example 3, whether feasible or not, could be found simply by designating four of the variables as non-basic and solving for the remaining five. Doing this for all possible combinations of four variables would yield every possible basic solution. The cost for each solution can then be evaluated and the solution with the lowest cost would be the optimal solution. This method has serious drawbacks as can be seen when considering a system with ten origins and ten destinations. For this system one would have to solve $C(100,19)$ sets of 81 equations in 81 unknowns. Even with the use of a high-speed computer this problem would take several weeks. Using the distributive methods of this section and VAM for the initial solution, a skilled operator could work the same problem by hand in 3 to 4 hours. Use of an IBM 1410 computer could reduce this time to about 25 minutes. As was indicated in the introduction, the Simplex method, also an iterative method, could be used. Its formulation of the problem would have 119 variables and 19 elements in each basis, and would require 30 to 40 minutes using an IBM 1410 computer to reach an optimal solution.

In the following two iterative methods for obtaining an optimal solution for the transportation problem, an initial solution is first obtained. This solution is then analyzed to see

if a better solution exists. If not, the present solution is considered to be optimal. If by some indication the present solution is not optimal, the method outlined in Theorem 4 is used to obtain a solution which is better than the first. This solution is then checked for optimality and the process is continued.

The Distribution method permits any size transportation problem to be systematically solved with the best solution resulting. Due to the special characteristics of the transportation matrix and the mathematics of the method, the final solution is exact rather than approximate. The Distribution method is really little more than a systematic inspection method and has been replaced, for the most part, by the Modified Distribution method (MOD). These distribution methods are best worked using the distribution matrix or chart form of representing the data rather than the simplex equation form.

For the initial solution the arbitrary Northwest Corner Rule will be used. In actual practice, however, almost any initial basic feasible solution may be used. In order to determine whether or not this is an optimal solution, that is, whether or not the solution variables included in the solution minimize the objective function, the effects on the total cost function of introducing one or more of the currently excluded variables must be determined.

Consider the solution in Fig. 13 again, which is the initial solution obtained by the Northwest Corner Rule for Example 3. For easy reference, the squares in which assignments appear will

be called S squares, with S_{ij} denoting the square in which the assignment x_{ij} appears. Similarly, W_{ij} will denote the square common to the i^{th} row and the j^{th} column and containing no assignment. Thus each W square is considered one at a time with an assignment of one unit made in it. Of course, the introduction of one unit in W_{ij} necessitates the modification of some of the existing assignments to keep the rim conditions satisfied. Theorem 4 is assurance that this modification can be made in only one way.

In this particular example, if one unit is shipped from O_2 to D_1 and introduced in W_{21} , the assignment made in S_{11} must be reduced by one unit to satisfy the rim requirement of the first column. Similarly, S_{12} must be increased by one unit and S_{22} decreased by one unit. The introduction of one unit in W_{21} produces the program in Fig. 23.

The intention here is to determine whether or not an introduction of x_{21} into the program will reduce the cost of the present program. If it will, then x_{21} will be allowed to increase until one of the current solution variables becomes zero, resulting in a new basic solution. The proportionality assumption is assurance that if the introduction of one unit in x_{21} will decrease the cost by c , then k units introduced will decrease the cost by kc . If this introduction of one unit in W_{21} does not indicate that the cost of the program will be decreased, then a new W square is considered.

The cost of the original solution in Fig. 13 was previously determined to be 1670. The cost of the new program with one

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		20		7		4	60
	49		11				
O_2		17		3		6	40
	1		39				
O_3		15		10		6	60
			30		30		
Req.		50		80		30	160

Fig. 23.

unit introduced in W_{21} is 1671. An additional cost, then, of +1 would be incurred for every unit of x_{21} put into the program, which indicates that no new assignment should be made in W_{21} . This additional cost of +1 is the opportunity cost for W_{21} , denoted O_{21} , and thus $O_{21} = +1$. Obviously any positive O_{ij} will indicate that an assignment in W_{ij} will not lead to a decrease in the cost function Z , and a new W square should be checked.

A similar argument shows that a negative opportunity cost associated with some W square indicates that any assignment in that square would reduce the total cost of the program. The occurrence of a zero opportunity cost indicates that an assignment made in that square would not change the cost of the program at all. Thus if each O_{ij} , for every W_{ij} , in a solution is nonnegative, the solution cannot be improved and is therefore

optimal.

A much easier way to determine O_{21} for W_{21} is the following. For each assignment which was increased by one unit around the path, Z is increased by its cost coefficient, and Z was decreased by each cost coefficient for each assignment which was decreased by one unit. Thus it is seen that $O_{21} = c_{21} - c_{11} + c_{12} - c_{22} = 17 - 20 + 7 - 3 = +1$. Since W_{21} indicates no possible improvement, a new W square is selected for consideration, say W_{31} . An assignment of one unit in W_{31} will necessarily reduce both S_{11} and S_{32} by one unit and increase S_{21} by one unit. Hence, $O_{31} = c_{31} - c_{11} + c_{12} - c_{32} = 15 - 20 + 7 - 10 = -8$. Since O_{31} is negative, this indicates that an assignment in W_{31} will bring about a decrease in Z . Thus x_{31} is increased until one of the current assignments on the path becomes zero. Since an increase in x_{31} decreases x_{11} and x_{32} , x_{32} will become zero when x_{31} is 30. An increase of more than thirty units would make x_{32} negative. At this point, $x_{11} = 20$ and $x_{21} = 40$. This, then, is a new basic solution, derived from the initial one, as shown in Fig. 24. The cost function for this new solution is

$$Z = 20 \times 20 + 7 \times 40 + 15 \times 30 + 6 \times 30 = 1430.$$

Rather than recompute Z for each new solution as was just done, a much more efficient method is to note that the introduction of thirty units in x_{31} reduced the cost of the original solution by $30 \times (+8) = 240$. The cost of this new program then is $1670 - 240 = 1430$.

The same process as was just described is now used on this

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁	20	20	40	7	30	4	60	60
O ₂		17	40	3		6	40	40
O ₃	30	15		10	30	6	60	60
Req.		50		80		30		160

Fig. 24.

first improved solution. Hopefully this solution will be optimal which would be indicated by a nonnegative opportunity cost evaluation for each W square. However, a check on W_{13} shows this is not the case. If x_{13} is increased by one unit, x_{11} and x_{33} are necessarily decreased by one unit, x_{31} increased by one, and

$$O_{13} = c_{13} - c_{33} + c_{31} - c_{11} = 4 - 6 + 15 - 20 = -7 .$$

Thus x_{31} is increased by 20 making $x_{11} = 0$, $x_{33} = 10$, and $x_{31} = 50$. This second improved solution has a cost $Z = 1430 - 7 \times 20 = 1290$ and is shown in Fig. 25 along with the opportunity costs associated with its W squares. Since all of the opportunity costs are positive, this solution is optimal.

Notice that to find O_{21} , if x_{21} is increased by one unit, x_{31} and x_{22} are decreased by one unit, x_{12} and x_{33} are then

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		20		7		4		60
			40		20			
O ₂		17		3		6		40
			40					
O ₃		15		10		6		60
	50				10			
Req.		50		80		30		160

$$\begin{aligned}
 O_{11} &= 7 & O_{32} &= 1 \\
 O_{21} &= 8 & O_{23} &= 6
 \end{aligned}$$

Fig. 25.

increased by one unit and x_{13} decreased by one. Thus

$$\begin{aligned}
 O_{21} &= c_{21} - c_{22} + c_{12} - c_{13} + c_{33} - c_{31} \\
 &= 17 - 3 + 7 - 4 + 6 - 15 = 8.
 \end{aligned}$$

All five assignments were affected in this case, whereas only three have been in the W square evaluations considered before. This exemplifies the fact that there is no specific method for establishing the path for each W square. Rules can be stated for computational purposes; however, these are nothing more than trial and error procedures.

Another point of interest is the number of iterations that were necessary to reach the optimal solution. In this example only two were necessary, and were made immediately upon finding

a W square which indicated improvement was possible. For larger problems in which the paths become much more complicated, this is probably the best procedure. However, evaluating all of the opportunity costs for each solution, and choosing the one which will decrease Z the most, will usually cut down on the number of iterations necessary to reach the optimal solution. For larger problems, most of the computation time is spent in searching for the path. Once found the new solution can be obtained almost immediately. For this reason the method used in the working of the example is probably more efficient than evaluating each W square for every new solution.

The basic parts of the Distribution method can be summarized in the following steps.

1. Establish an initial solution.
2. Evaluate the W squares one by one until one is found which shows improvement.
 - a. Establish a closed path from this selected W square via S squares back to the same W square. Other S and W squares may be skipped over.
 - b. Determine the improvement possibility for this W square by calculating its opportunity cost.
 - c. If no possible improvement is indicated repeat step 2 with a different W square.
 - d. When a W square is found which shows improvement possibility, assign as much as possible to this square keeping all rim requirements satisfied and

bringing one of the current assignments to zero.

3. Repeat step 2 until no further improvement is possible.

The solution is then optimal.

Concerning the merits of this method, it can be said that it does yield an optimal solution and does work toward it in a straightforward manner. However, it is at best little more than a formalized inspection method. In larger problems the method becomes tedious if worked by hand and much of the computation time is spent searching for the proper path. This method has been replaced for the most part by the more efficient Modified Distribution method.

The Modified Distribution method is similar to the Distribution method in all respects except one. For each W square in the previous method, a closed path through some or all of the S squares had to be found. Then the opportunity cost of each W square could be evaluated. Essentially the Modified Distribution method evaluates these opportunity costs before any paths are found. The W square with the greatest possibility of improvement is located and then the corresponding path determined. The important thing is that in this method, only one path need be found per iteration.

A complete set of row and column numbers is found such that the cost coefficient in each S square equals the sum of its row and column numbers. Using Example 3 again, with the initial solution obtained by the Northwest Corner Rule, the program in chart form with the addition of the extra row and column is given in Fig. 26.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply	⋮
		20		7		4			
O ₁	50		10					60	R ₁
		17		3		6			
O ₂			40					40	R ₂
		15		10		6			
O ₃			30		30			60	R ₃
Req.	50		80			30		160	
		K ₁		K ₂		K ₃			

Fig. 26.

For each S square, numbers R_i and K_j are to be found such that $R_i + K_j = c_{ij}$. Thus the following equations result:

$$\begin{aligned}
 c_{11} &= R_1 + K_1 \\
 c_{12} &= R_1 + K_2 \\
 c_{22} &= R_2 + K_2 \\
 c_{32} &= R_3 + K_2 \\
 c_{33} &= R_3 + K_3
 \end{aligned}
 \tag{7}$$

Since the c 's are known, this is seen to be a system of 5 equations in 6 unknowns. To evaluate the R 's and K 's, one of them is arbitrarily fixed, and for computational purposes the simplest thing to do is set $R_1 = 0$. Figure 27 gives the results of solving the system of equations for the remaining R_i

and K_j values.

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply	\vdots
O_1	50	20	10	7		4		60	0
O_2		17	40	3		6		10	-4
O_3		15	30	10	30	6		60	3
Req.		50		80		30		160	
		20		7		3			

Fig. 27.

With these constants evaluated, the opportunity costs for each W square are given by $O_{ij} = -(R_i + K_j) + c_{ij}$. This is certainly not immediately evident. The following argument should suffice, however, to indicate why it is so.

On solving (7) for all of the K 's and R 's, with the exception of R_1 , in terms of the c 's and R_1 , notice that R_1 appears in each equation for K_j with a coefficient of -1 , and in each equation for R_i with a coefficient of $+1$. Thus adding some R_i and K_j together gives an expression involving only c 's no matter what value is given to R_1 initially.

$$\begin{aligned}
 K_1 &= c_{11} - R_1 \\
 K_2 &= c_{12} - R_1 \\
 R_2 &= c_{22} - c_{12} + R_1 \\
 R_3 &= c_{32} - c_{12} + R_1 \\
 K_3 &= c_{33} - c_{32} + c_{12} - R_1
 \end{aligned}
 \tag{8}$$

In the Distribution method, the path found for W_{21} determined the opportunity cost $O_{21} = c_{21} - c_{11} + c_{12} - c_{22}$. From (8),

$$\begin{aligned}
 -(R_2 + K_1) &= -c_{22} + c_{12} - R_1 - c_{11} + R_1 \\
 &= -c_{22} + c_{12} - c_{11}
 \end{aligned}$$

and thus

$$O_{21} = c_{21} - (R_2 + K_1) = c_{21} - c_{22} + c_{12} - c_{11}$$

which is the same determination of O_{12} as was found using the Distribution method. A more rigorous argument would involve the technical implications of the definition of opportunity cost.

The opportunity costs given by $O_{ij} = c_{ij} - (R_i + K_j)$ for each W square are now computed.

$$\begin{aligned}
 O_{13} &= c_{13} - (R_1 + K_3) = 4 - (0 + 3) = 1 \\
 O_{21} &= c_{21} - (R_2 + K_1) = 17 - (-4 + 20) = 1 \\
 O_{23} &= c_{23} - (R_2 + K_3) = 6 - (-4 + 3) = 7 \\
 O_{31} &= c_{31} - (R_3 + K_1) = 15 - (3 + 20) = -8
 \end{aligned}$$

Since O_{31} is negative, this indicates that an assignment in x_{21} will decrease the total cost function just as in the Distribution method. In fact, this was the assignment made to arrive at the first improved solution for Example 3 in Fig. 24 using the Distribution method.

Since the S squares are no longer the same, the R's and K's must be recalculated.

$$c_{11} = R_1 + K_1$$

$$c_{12} = R_1 + K_2$$

$$c_{22} = R_2 + K_2$$

$$c_{31} = R_3 + K_1$$

$$c_{33} = R_3 + K_3$$

Setting $R_1 = 0$ again and solving for the other R's and K's gives

$$R_1 = 0 \qquad K_1 = 20$$

$$R_2 = -4 \qquad K_2 = 7$$

$$R_3 = -5 \qquad K_3 = 11$$

$$O_{13} = c_{13} - (R_1 + K_3) = 4 - (0 + 11) = -7$$

$$O_{21} = c_{21} - (R_2 + K_1) = 17 - (-4 + 20) = 1$$

$$O_{23} = c_{23} - (R_2 + K_3) = 6 - (-4 + 11) = -1$$

$$O_{32} = c_{32} - (R_3 + K_2) = 10 - (-5 + 7) = 8$$

Since $O_{13} = -7$, the assignment is to be made in x_{13} . The appropriate path is found and the new solution given in Fig. 28 is determined.

This was the optimal assignment obtained by the Distribution method. Checking to see that it is, the R's and K's are computed once more.

$$c_{12} = R_1 + K_2$$

$$c_{13} = R_1 + K_3$$

$$c_{22} = R_2 + K_2$$

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		20		7		4		60
			40		20			
O ₂		17		3		6		40
			40					
O ₃		15		10		6		60
	50				10			
Req.		50		80		30		160

Fig. 28.

$$c_{31} = R_3 + K_1$$

$$c_{33} = R_3 + K_3$$

Again setting $R_1 = 0$, and solving for the values of the rest of the R's and K's, gives

$$R_1 = 0 \qquad K_1 = 13$$

$$R_2 = -4 \qquad K_2 = 7$$

$$R_3 = 2 \qquad K_3 = 4$$

The evaluation of the W squares

$$O_{11} = c_{11} - (R_1 + K_1) = 20 - (0 + 13) = 7$$

$$O_{21} = c_{21} - (R_2 + K_1) = 17 - (-4 + 13) = 8$$

$$O_{23} = c_{23} - (R_2 + K_3) = 6 - (-4 + 4) = 6$$

$$O_{32} = c_{32} - (R_3 + K_2) = 10 - (2 + 7) = 1$$

is identical with that obtained following the last iteration in the Distribution method.

Although for this particular example there were two iterations necessary to reach the optimal solution for each of the methods, the Modified Distribution method required the finding of only two paths, whereas the Distribution method required the finding of four just to check the optimality of the last solution. On first thought, it would seem that the solving of the equations for the R and K values would require a good deal of time. This is not the case, however, since setting $R_1 = 0$ immediately determines some K_j , which in turn determines some other R_i and so on. In fact, this procedure is easily done by mere inspection.

DEGENERACY

Although much has already been said about this topic, a more complete discussion is included here to justify some of the work done in the next sections and to describe a particularly easy method for dealing with degenerate solutions. Example 3 was chosen to illustrate the two methods just presented because it was known to have no possibility of degenerate solutions. More often than not, however, degeneracy is found to occur in some solution when applying the previous two methods to solve a transportation problem.

Consider the following example in Fig. 29 in which the rim conditions of Example 3 have been modified slightly, and the

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
O_1		20		7		4		50
	50							
O_2		17		3		6		50
			50					
O_3		15		10		6		60
			30		30			
Req.		50		80		30		160

Fig. 29.

Northwest Corner Rule used to obtain an initial solution. Notice that only four assignments were made, one less than the required number for a basic feasible solution. As was said before, this indicates degeneracy has occurred. The first assignment in using the Northwest Corner Rule was made to x_{11} . An assignment of 50 to x_{11} simultaneously satisfied both of the requirements of the first row and the first column, a situation which was assumed impossible in the construction of the algorithm.

One method to resolve degeneracy would be to modify the algorithm. The case where an assignment, other than the last one, simultaneously satisfies both a row and column requirement was originally said to be impossible. An additional step such as the following can be added which will take care of this case. If an assignment simultaneously satisfies both a row and column

requirement, delete either the row or the column, but not both. Looking now at the example, the assignment of $x_{11} = 50$ reduces both a_1 and b_1 to zero. Assume now the row is ignored, x_{21} becomes the new northwest corner element, and $x_{21} = \text{Min}(a_2, b_1) = \text{min}(50, 0) = 0$. Entering this assignment as a zero will distinguish it from the other zero assignments which are simply left blank. If the interpretation is made that this zero assignment in x_{21} can be actually thought of as a small positive number very close to zero, the solution will now have the required five solution variables, making it thus a basic feasible solution. Distinguishing a zero assignment as an S square and a blank assignment as a W square, enables one to easily cope with the problem of degeneracy. This is a much more efficient method of dealing with this problem than the introduction of a perturbation constant when using hand computations.

The above example was a case where degeneracy occurred in the initial solution. The method presented to resolve it may be used at any stage in the algorithm. If in some later iteration degeneracy occurs, it can be detected as before by noting the lack of a sufficient number of solution variables. There are other ways of detecting its occurrence. In the Distribution method, the failure to find a path for a particular W square indicates that not enough S squares were obtained in the last iteration, and hence less than the correct number of assignments was made to obtain a basic feasible solution. In the Modified Distribution method, a system of too few equations will result in attempting to establish the R and K values.

A method much more mathematically pleasing and computer orientated, of resolving degeneracy, is the perturbation method. To each of the column requirements is added a small quantity ϵ and to the last row requirement the quantity $n\epsilon$. This ϵ is an arbitrarily small number which will not affect the accuracy of the solution. For computer use, five place accuracy might be sufficient for the answers. If the work is done to eight place accuracy with an $\epsilon = 1.0 \times 10^{-8}$, when the answers are rounded to five places the ϵ has no effect. The nicest feature of this method is that the rim requirements can be changed at the beginning of the algorithm and forgotten about. This would indicate that perturbation is probably the best method for resolving degeneracy when using a computer. At every point in the program where a zero assignment should be made to make the solution a basic one, it will appear as some multiple of ϵ . The previous problem has been worked completely in Fig. 30, using the North-west Corner Rule and the Distribution method for obtaining the optimal answer.

If ϵ were set equal to zero, Fig. 31 would give the same solutions obtained by using the first method to handle degenerate solutions. Note that when ϵ is set equal to zero, the first improved solution is the same as the initial solution. The assignment of zero units to x_{21} was moved to x_{31} . It is usually the case that a degenerate solution will involve extra iterations.

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
O_1	50	20		7		4		50
O_2	ϵ	17		$50-\epsilon$	3	6		50
O_3		15		$30+2\epsilon$	10	$30+\epsilon$	6	$60+3\epsilon$
Req.	$50+\epsilon$			$80+\epsilon$		$30+\epsilon$		$160+3\epsilon$
O_1	50	20		7		4		50
O_2		17		50	3	6		50
O_3	ϵ	15		$30+\epsilon$	10	$30+\epsilon$	6	$60+3\epsilon$
Req.	$50+\epsilon$			$80+\epsilon$		$30+\epsilon$		$160+3\epsilon$
O_1	$20-\epsilon$	20		$30+\epsilon$	7		4	50
O_2		17		50	3		6	50
O_3	$30+2\epsilon$	15			10	$30+\epsilon$	6	$60+3\epsilon$
Req.	$50+\epsilon$			$80+\epsilon$		$30+\epsilon$		$160+3\epsilon$
O_1		20		$30+\epsilon$	7	$20-\epsilon$	4	50
O_2		17		50	3		6	50
O_3	$50+\epsilon$	15			10	$10+2\epsilon$	6	$60+3\epsilon$
Req.	$50+\epsilon$			$80+\epsilon$		$30+\epsilon$		$160+3\epsilon$

Fig. 30.

UNEQUAL SUPPLY AND DEMAND

To build up the mathematics necessary to derive algorithms to work the transportation problem, it was necessary to add the assumption that supply and demand were equal. This is seldom the case, however, in a real problem. It should suffice to treat the case where the supply is greater than the demand. Consider Example 3 again, with the rim requirement in the third column reduced to 15 as shown in Fig. 31. There are now 15 more units

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots	Supply
O_1		20		7		4		60
O_2		17		3		6		40
O_3		15		10		6		60
Req.		50		80		15		

Fig. 31.

available at the origins than are required by the destinations. A simple way to formulate this problem as one in which the supply is equal to the demand is to add a "dummy" destination which will receive these 15 extra units of goods. Since these goods will not actually be shipped to a destination and will remain at one or more of the origins, the cost to ship one unit of these goods

from each O_i to the dummy destination will be zero. The example is worked in Fig. 32 using VAM for the initial solution, and the Distribution method to obtain the optimal solution. Here O_1 is left with five units unshipped and O_3 with ten.

	D_1	D_2	D_3	Dummy	Supply
O_1	20	7	4	0	60
	5	40	15		
O_2	17	3	6	0	40
		40			
O_3	15	10	6	0	60
	45			15	
Req.	30	80	15	15	160
O_1	20	7	4	0	60
		40	15	5	
O_2	17	3	6	0	40
		40			
O_3	15	10	6	0	60
	50			10	
Req.	30	80	15	15	160

Fig. 32.

OTHER RESTRICTIONS

In the problems considered so far any shipping route could be used. The previous example, when solved, indicated the best solution in terms of least cost included nonshipment of five units from O_1 and ten from O_3 . It is possible to add various restrictions to the problem which can be handled very conveniently by distributive methods. The problems considered to this point so far have minimized total cost. Hence any shipping route involving high costs has been avoided. Consider now what would happen if O_3 had to get rid of all of its supply to assure facilities for incoming products. The above program is no longer workable. The clue to handling conditions such as this is to introduce a very high cost coefficient for any route to be avoided. Originally the cost to retain one unit of goods at O_3 was zero. If this cost is raised to some arbitrarily large number, the optimizing techniques will tend to avoid this assignment. To assure this cost is sufficiently high, it is entered as M , defined simply as arbitrarily large enough to avoid the assignment. For computer operation, the choice of M as at least 100 times the maximum of the other costs should suffice.

Assume now that O_3 must get rid of all its supply. Using VAM the solution is optimal the first time and appears in Fig.33. If the problem were first formulated in this way the user would not find out what the optimal solution could be. Thus something can be said for obtaining the optimal solution for the unrestricted problem and evaluating what changes and costs are

	∴ D ₁	∴ D ₂	∴ D ₃	∴ Dummy	∴ Supply
O ₁	20	7	4	0	60
		40	5	15	
O ₂	17	3	6	0	40
		40			
O ₃	15	10	6	M	60
	50		10		
Req.	50	80	15	15	160

Fig. 33.

involved to reach a solution compatible with the restrictions.

ALTERNATE SOLUTIONS

In some cases, when an optimal solution is reached, there exist other solutions which are also optimal. As of now nothing has been said regarding these other possibilities. For the user to make decisions regarding shipping program possibilities, more information is needed. It may be the case that one of these other optimal solutions, or even one which is not optimal but close to optimal, is a better policy to follow than the one arrived at. The final tableau leaves much to be desired as far as "extra" information is concerned. It gives only the assignments which will yield a minimum total cost and nothing more. The

problem then is to find out what extra information, if any, can be gleaned from the algorithm once an optimal solution has been found.

Alternate solutions can be obtained by exactly the same procedure as used to obtain a new solution. The first step is to make a check on the W squares. This has already been done when checking to see if the last iteration yielded an optimal solution. None of the W squares were negative, indicating the solution was optimal. However, if one or more of the opportunity costs at this point is zero, this indicates that other solutions, as well as the one obtained, have the same minimal cost figure, and thus are themselves optimal. The path is found for each of these squares and assignments made in them. Any positive assignment consistent with the rim requirement will then yield an optimal solution as will be seen by the following example. The costs in Example 3 have been slightly modified and the problem reworked. The optimal solution and the opportunity costs for the W squares are given in Fig. 34. The cost associated with this solution is 1290. As has been said, the zero opportunity cost for W_{32} indicates that certain assignments can be made in W_{32} which will also yield an optimal solution. This makes sense in that an assignment in this square will neither increase nor decrease the total cost of the program. Moving all that is possible into this allocation gives the solution in Fig. 35 having a total cost also of 1290. Notice that any part of the ten units that were moved could also have been the allocation made in W_{32} . Figure 36 gives the solution obtained by making an

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		14		7		4		60
			40		20			
O ₂		17		3		6		40
			40					
O ₃	50	15		9		6		60
					10			
Req.		50		80		30		160

$$O_{11} = 1 \qquad O_{32} = 0$$

$$O_{12} = 8 \qquad O_{23} = 6$$

Fig. 34.

	⋮	D ₁	⋮	D ₂	⋮	D ₃	⋮	Supply
O ₁		14		7		4		60
			30		30			
O ₂		17		3		6		40
			40					
O ₃	50	15		7		6		60
			10					
Req.		50		80		30		160

Fig. 35.

	⋮	D_1	⋮	D_2	⋮	D_3	⋮	Supply
O_1		14		7		4		60
			34		26			
O_2		17		3		6		40
			40					
O_3		15		9		6		60
	50		6		4			
Req.		50		80		30		160

Fig. 36.

assignment of six units in x_{32} . The cost of this solution is also 1290.

Assuming that a unit is indivisible, this means that there are eleven optimal solutions for this particular problem. Although it seems that making two deliveries to one location where one would suffice would tend to raise costs somewhat, any one of these solutions would be acceptable. The last chart in which six allocations were made, instead of the necessary five to have a basic feasible solution, does not violate the statement in Theorem 4 that optimal solutions are basic feasible solutions. This simply indicates that optimal solutions can exist which are not basic ones. Theorem 4 only guarantees that no nonbasic solution exists which is better than a basic optimal one.

It may be possible, for some reason or another, that the

user wishes to know about some of the near optimal solutions which also exist. An analysis of these can be made in a similar manner. Looking again at Fig. 34, an introduction of one unit into W_{11} will increase the cost of the program by 1. Again this is indicated by the W_{11} square evaluation obtained in checking the optimality of the solution. This program, in Fig. 37, would have a total cost of 1291. The introduction of one unit into

	\vdots	D_1	\vdots	D_2	\vdots	D_3	\vdots Supply
O_1		14		7		4	60
	1		40		19		
O_2		17		3		6	40
			40				
O_3		15		9		6	60
	49				11		
Req.		50		80		30	160

Fig. 37.

W_{21} would increase the objective function to 1298. In a similar manner the opportunity costs of the other W squares can give additional information. This analysis is useful in determining "next best" solutions which are near optimal. This particular analysis would indicate that if an optimal solution is for some reason unsatisfactory, an allocation in W_{11} would be more

beneficial than one in W_{21} .

There are other questions which can be answered based on this optimal solution. For instance, what range in values can c_{11} take on and still leave this solution optimal? If c_{11} is reduced from 14 to 12, the solution is no longer optimal since the opportunity cost for this square now becomes -1. Thus if a cheaper way can be found to ship goods along this route the present program should probably be re-evaluated. There will be a lower limit on the cost coefficient for each W square for which the program is optimal. A similar analysis can be made of each S square cost coefficient. Consider c_{31} in the previous example. A change in its value affects the opportunity costs of W_{11} and W_{12} . If c_{31} is increased from 15 to 17, then the opportunity cost for W_{11} becomes -1, and again indicates the program can now be improved. A decrease in c_{31} only reduces the cost of the program and makes it more "stable" in the sense that it is now less sensitive to a change. Thus the cost coefficient in each square has an upper limit which, if exceeded, will indicate the solution is no longer optimal.

Other analyses on things such as changes in rim requirements seem interesting but are left untouched by most authors. For instance, in the previous example, if D_3 decreases its order by 10 units, which origin should take up the slack? Both O_1 and O_3 supply 10 units to D_3 , but O_1 supplies them at a greater cost. It would seem to make sense to cut the supply from O_1 to D_3 , rather than from O_3 to D_3 . However, in doing this and leaving 10 units at O_1 unassigned, the program is no longer optimal.

A better solution is possible, namely, cutting the supply from O_3 which will not reduce the cost as much.

CONCLUSION

Although the distributive methods for solving the transportation problem are actually little more than formalized inspection methods, they are fast, exact, and easy to use. For small problems they are excellent. Their application to large problems usually requires the use of a computer as do nearly all linear programming methods.

The various merits of the distributive methods have been mentioned throughout this paper. Although they are faster than most methods in reaching an optimal solution, they are still very inefficient. Approximately 50 per cent of the user's computational time is spent searching for the proper paths when using the Modified Distribution method. Even so, the distributive methods seem to be the most popular linear programming methods for solving transportation problems.

ACKNOWLEDGMENT

The author wishes to express his thanks and appreciation to Professor S. T. Parker for his helpful suggestions, comments, and careful checking of this report.

BIBLIOGRAPHY

1. Charnes, Abraham and W. W. Cooper.
Management Models and Industrial Applications of Linear Programming. New York: John Wiley and Sons, Inc., 1961.
2. Dantzig, George Bernard.
Linear Programming and Extensions. Princeton: University Press, 1963.
3. Garvin, Walter W.
Introduction to Linear Programming. New York: McGraw-Hill Book Company, Inc., 1960.
4. Gass, Saul I.
Linear Programming. New York: McGraw-Hill Book Company, 1964.
5. Horowitz, Ira.
An Introduction of Quantitative Business Analysis. New York: McGraw-Hill Book Company, 1965.
6. Llewellyn, Robert W.
Linear Programming. New York: Holt, Rinehart and Winston, 1964.
7. Lomba, Paul N.
Linear Programming. New York: McGraw-Hill Book Company, 1964.
8. Metzger, Robert W.
Elementary Mathematical Programming. New York: John Wiley and Sons, Inc., 1963.
9. Reinfield, Nyles V. and William R. Vogel.
Mathematical Programming. Englewood Cliffs: Prentice-Hall, Inc., 1958.

THE TRANSPORTATION PROBLEM

by

DALE GARRATT SHAW

B. A., Colorado College, 1963

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Mathematics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

A particularly interesting special case of the general linear programming problem is the transportation problem. This paper contains a detailed discussion of so-called distributive methods for obtaining a solution for any given transportation problem. The transportation theory underlying these methods is developed entirely independently of the programming theory of the simplex method. The transportation problem is first stated with the added restrictions necessary to construct a mathematical model. Definitions are made and four fundamental theorems of transportation theory are stated and proved.

The Distribution method, the Modified Distribution method, and various methods for constructing initial solutions are presented in detail and indications as to the efficacy of these methods and others are given. Problems concerning degeneracy, unequal supply and demand, restricted routes, and certain added conditions are considered and shown to be easily resolved using distributive methods.

Finally, methods are given for finding alternate solutions, concluding with a brief discussion of the area of sensitivity analysis.