

605 056  
605056

✓

HITCHCOCK TRANSPORTATION PROBLEM  
D. R. Fulkerson  
✓ P-890 PH  
July 9, 1956

COPY	1	OF	1	30-P
HARD COPY				\$ .200
MICROFICHE				\$ .050

Approved for OTS release

DDC  
 SEP 9 1964  
 The RAND Corporation  
 1700 MAIN ST. • SANTA MONICA • CALIFORNIA  
 DDC RA E

**Best  
Available  
Copy**

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI ,  
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # *11605056*

- 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.
- 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.
- 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.
- 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.
- 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.
- 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.
- 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.
- 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI ( TT DOCUMENTS ONLY).
- 9.

NBS 9 64

PROCESSOR: *117*

SUMMARY

↓  
An exposition of the simplex computation for transportation  
type problems. ( ↘

CONTENTS

SUMMARY .....	11
Section	
1. SIMPLEX METHOD .....	6
2. OPTIMAL ASSIGNMENT PROBLEM .....	19
3. UPPER BOUNDS ON VARIABLES .....	20
4. THE TRANSSHIPMENT PROBLEM .....	21
REFERENCES .....	26

## HITCHCOCK TRANSPORTATION PROBLEM

D. R. Fulkerson

The transportation problem was first formulated by F. L. Hitchcock [12] in 1941; he also gave a computational procedure, much akin to the general simplex method, for solving the problem. Independently, during World War II, T. C. Koopmans arrived at the same problem in connection with his work as a member of the Joint Shipping Board. The problem is thus frequently referred to as the Hitchcock-Koopmans problem.

Mathematically the problem has the form

$$(1) \quad \text{minimize} \quad \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ij}$$

subject to the constraints

$$(2a) \quad x_{ij} \geq 0$$

$$(2b) \quad \sum_j x_{ij} = a_i \quad (i = 1, \dots, m)$$

$$(2c) \quad \sum_i x_{ij} = b_j \quad (j = 1, \dots, n)$$

where  $a_i$ ,  $b_j$ ,  $a_{ij}$  are given. Feasibility is assured by assuming  $a_i \geq 0$ ,  $b_j \geq 0$ , and  $\sum_i a_i = \sum_j b_j$ . A particular realization of the problem appears if we think of  $m$  sources of a commodity, the  $i^{\text{th}}$  one possessing  $a_i$  units, and  $n$  sinks, the  $j^{\text{th}}$  one requiring  $b_j$  units, and interpret  $a_{ij}$  as the unit transportation cost from source  $i$  to sink  $j$ . Thus the linear form (1) becomes the total

transportation bill, to be minimized subject to fulfilling demands at the sinks from supplies at the sources. This interpretation is a paraphrase of Hitchcock's original statement of the problem.

To give some idea of the diverse applications of the transportation problem, we have compiled the following list (by no means exhaustive) of problems which can be successfully attacked, either directly as transportation problems or by subsequent developments of the theory:

(1) Suppose  $n$  men are to be assigned to  $n$  jobs, where man  $i$  in job  $j$  has a "score"  $a_{ij}$ ; (a) find an assignment of men to jobs which maximizes the sum of the scores [19]; (b) find an assignment which maximizes the least score [9].

(2) Consider a network consisting of  $N$  points and interconnecting links. Let  $a_i \geq 0$  denote the supply of a commodity at the  $i^{\text{th}}$  point,  $b_i \geq 0$  the demand at the point, where  $a_i b_i = 0$ , and  $a_{ij} \geq 0$  the unit cost of shipping from  $i$  to  $j$ . Assuming that  $\sum a_i = \sum b_i$ , and that any point can act as a transshipment point, find a minimal cost transportation program [16].

(3) Given a network, let  $a_{ij}$  denote the length of the link from  $i$  to  $j$ . Find a chain of minimal length connecting two given points [16].

(4) Suppose  $B = (b_{ij})$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ , is a given matrix. By a "walk through  $B$  from  $m_n$  to  $1_n$ " we will mean a collection of elements  $b_{ij}$  which can be arranged as

$$b_{mn}, b_{1_1j_1}, \dots, b_{1_kj_k}, b_{1n}$$

where each element occurs (in B) either directly above or to the right of the preceding one. If the sum of the displayed members is the length of the walk, find a walk of minimal length [6].

(5) Let  $t_{1j}^k$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ,  $k = 1, 2, \dots$ , be times at which carriers are to pick up loads at origin  $i$  for delivery to destination  $j$ . Assuming we know travel times between each origin-destination pair, what is the minimum number of carriers required to meet the schedule [3]?

(6) Suppose given a network of  $N$  points, where some point acts as a source for a commodity, another as a sink, and the remaining as transshipment points. If each link of the network has a nonnegative number assigned to it, its flow capacity, find a maximal flow from source to sink [4, 5, 7, 8].

(7) Given a capacitated network (as in (6)), let  $a_i \geq 0$  ( $b_i \geq 0$ ) denote the supply (demand) at the  $i^{\text{th}}$  point. Under what conditions is it possible to fulfill the demands from the supplies [10]? Find a minimal cost transportation program.

(8) Let  $E_1, \dots, E_n$  be subsets of a given finite set  $E = \{e_1, \dots, e_m\}$ . Under what conditions is it possible to pick out distinct  $e_1, \dots, e_n$  such that  $e_j \in E_j$ ,  $j = 1, \dots, n$  [10, 11]?



(9) Let  $P$  be a finite partially ordered set consisting of elements  $p_1, \dots, p_N$ . A chain of  $P$  is a set of one or more distinct  $p_{1_1}, \dots, p_{1_k}$  with  $p_{1_1} \geq p_{1_2} \geq \dots \geq p_{1_k}$ . A set of disjoint chains covers  $P$  if each element of  $P$  belongs to some chain of the set. Find a minimal covering of  $P$  [13].

(10) A caterer knows that he will need  $r_j > 0$  fresh napkins on the  $j^{\text{th}}$  day,  $j = 1, \dots, n$ . Laundering normally takes  $p$  days; however the laundry has a higher cost service taking  $q < p$  days. Laundering costs  $b$  and  $c$  cents respectively, and new napkins  $a$  cents,  $b < c < a$ , say. How does the caterer meet his needs at minimal cost [14, 17]?

(11) Given a capacitated network (as in (6)) with source and sink, suppose it takes one unit of time to ship from any point to a neighboring point. Maximize the total flow through the network from source to sink in the first  $N$  time periods.

(12) Given a warehouse with fixed capacity,  $b$  and an initial stock  $s$  of a certain commodity, which is subject to known seasonal price and cost variations, what is the optimal pattern of purchasing, storage, and sales [1]?

Perhaps more impressive, from a practical viewpoint, than a long list of transportation-type problems is a recent report [18] from a panel discussion on "The Current Uses of Linear Programming in American Business" which states that out of about twenty examples of the industrial uses of linear programming known to the panel, more than half of these were transportation problems. It is consequently important to find

more efficient ways to compute such problems (even though the simplex computation to be discussed presently is certainly efficient), particularly since larger and larger systems are being considered. (One practical example reported lately has  $m = 200$ ,  $n = 3000$ .)

From a mathematical point of view, perhaps the most interesting distinguishing feature of the transportation problem is that it provides an approach to some problems which at first appear to be purely combinatorial. Thus, for example, problems 1, 3, 4, 5, 8, 9, 10 in the list are of this nature. The reason for this, as we shall see, is that in case the  $a_i$  and  $b_j$  are integral, the extreme points of the convex polyhedron described by (2) are also integral. Stated another way, this simply means that fractions are never encountered in solving the problem via the simplex method. This characteristic, which makes it possible to represent some combinatorial problems as continuous ones, also accounts for the ease of computation of transportation problems. While it is an impossible task, at present, to solve a general linear programming problem with 3200 equations and 600,000 variables, transportation problems of this size are currently being programmed. Even by hand, problems of fairly respectable size can be computed in a few hours. For example, L. R. Ford and I recently solved (using a new algorithm to be discussed at the end of the course) a problem with  $m = 12$ ,  $n = 120$ , involving all 1440 variables, in something like eight hours.

## 1. SIMPLEX METHOD

The particularization of the simplex method to the transportation problem was set forth by G. Dantzig in a paper published in 1951 [2]. Since then, various other accounts of the method have appeared and also other proposals for solving the problem have been made. It is certainly safe to assert that the simplex computation is the one in widest use at present.

Before taking up the computation via the simplex method, we shall first state and prove some fundamental properties of the problem.

Lemma 1. The problem is feasible if and only if  $a_i \geq 0$ ,  $b_j \geq 0$ , and  $\sum a_i = \sum b_j$ .

Proof. If the condition holds, then  $x_{ij} = \frac{a_i b_j}{\sum a_i}$  satisfy the constraints.

Clearly the conditions  $a_i \geq 0$ ,  $b_j \geq 0$  are necessary. Summing equations (2b) on  $i$ , (2c) on  $j$  yields  $\sum_i a_i = \sum_{i,j} x_{ij} = \sum_j b_j$ .

The  $m + n$  by  $mn$  coefficient matrix  $H$  of the set of equations (2) has the form

		$x_{11}$	$x_{12}$	$\dots$	$x_{1n}$	$x_{21}$	$x_{22}$	$\dots$	$x_{2n}$	$\dots$	$x_{m1}$	$x_{m2}$	$\dots$	$x_{mn}$
$m$	$H =$	1	1	$\dots$	1	0	0	$\dots$	0	$\dots$	0	0	$\dots$	0
		0	0	$\dots$	0	1	1	$\dots$	1	$\dots$	0	0	$\dots$	0
		$\cdot$	$\cdot$	$\dots$	$\cdot$	$\cdot$	$\cdot$	$\dots$	$\cdot$	$\dots$	$\cdot$	$\cdot$	$\dots$	$\cdot$
		0	0	$\dots$	0	0	0	$\dots$	0	$\dots$	1	1	$\dots$	1
$n$	$H =$	1	0	$\dots$	0	1	0	$\dots$	0	$\dots$	1	0	$\dots$	0
		0	1	$\dots$	0	0	1	$\dots$	0	$\dots$	0	1	$\dots$	0
		$\cdot$	$\cdot$	$\dots$	$\cdot$	$\cdot$	$\cdot$	$\dots$	$\cdot$	$\dots$	$\cdot$	$\cdot$	$\dots$	$\cdot$
		0	0	$\dots$	1	0	0	$\dots$	1	$\dots$	0	0	$\dots$	1

where we have recorded the variables  $x_{ij}$  at the top of their corresponding column vectors  $X_{ij}$ . If we refer to the first  $m$  components of  $X_{ij}$  as its source components, the last  $n$  as its sink components, then  $X_{ij}$  has 1 in its  $i^{\text{th}}$  source component, 1 in its  $j^{\text{th}}$  sink component, and zeros elsewhere.

Lemma 2. The matrix  $H$  has rank  $m + n - 1$ .

Proof. As we saw in the proof of Lemma 1, the sum of the first  $m$  rows of  $H$  is equal to the sum of the last  $n$ , and hence  $\text{rank } H \leq m + n - 1$ . On the other hand, it is easy to see that the  $m + n - 1$  column vectors  $X_{11}, X_{12}, \dots, X_{1n}, X_{2n}, \dots, X_{mn}$  are linearly independent. For assume a linear dependence

$$k_{11}X_{11} + k_{12}X_{12} + \dots + k_{1n}X_{1n} + k_{2n}X_{2n} + \dots + k_{mn}X_{mn} = 0.$$

This vector has the form (written as a row for convenience)

$$\left( \sum_{j=1}^n k_{1j}, k_{2n}, \dots, k_{mn}, k_{11}, k_{12}, \dots, k_{1,n-1}, \sum_{i=1}^m k_{ij} \right)$$

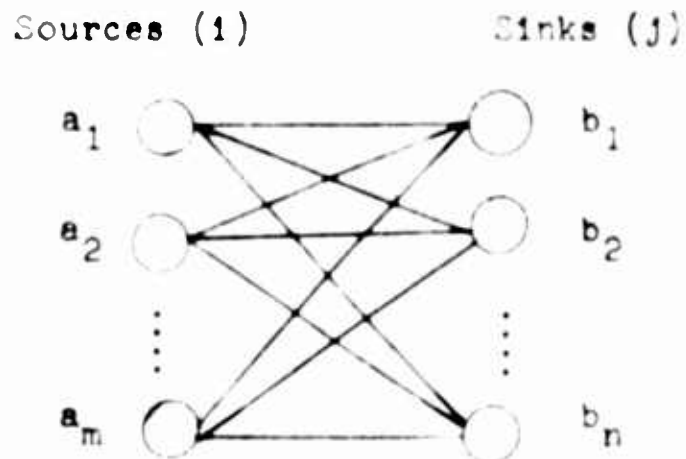
and consequently each  $k$  vanishes.

Thus one of the set (2b), (2c) could be discarded, leaving  $m + n - 1$  linearly independent equations. For computational purposes, it is perhaps better to retain all of the equations, however, and we shall do this.

There are various convenient ways of schematizing equations (2b), (2c). We might, for example, think of the array

$$\begin{array}{cccc}
 \left[ \begin{array}{cccc}
 x_{11} & x_{12} & \cdots & x_{1n} \\
 x_{21} & x_{22} & \cdots & x_{2n} \\
 \cdot & \cdot & \cdots & \cdot \\
 x_{m1} & x_{m2} & \cdots & x_{mn}
 \end{array} \right] & \begin{array}{l} a_1 \\ a_2 \\ \cdot \\ a_m \end{array} \\
 \begin{array}{ccc} b_1 & b_2 & b_n \end{array} & 
 \end{array}$$

having prescribed row and column sums  $a_i$  and  $b_j$ . Another way, that of the original realization of the problem, is to picture the linear graph consisting of  $m$  source-nodes,  $n$  sink-nodes, and all links  $ij$  joining sources to sinks



so that, associating  $x_{ij}$  with link  $ij$ , the  $i^{\text{th}}$  source equation says that the sum of all variables associated with links ema-

nating from  $i$  is  $a_i$ , and similarly for the sinks.

For hand computation, a layout sheet in the form of an array is commonly used and is more convenient than drawing linear graphs. For the purpose of proving theorems about the problem, however, we shall adopt the linear graph point of view for two reasons: to take advantage of some of the elementary concepts and results from linear graph theory, and to make the discussion of the Hitchcock problem tie in more closely with the next lecture on network problems.

Notice that the matrix  $H$  is the incidence matrix of nodes vs. links of the graph, that is, if we list the links  $ij$  horizontally, the nodes vertically, and record a 1 if a link is incident with a node, zeros elsewhere, the matrix  $H$  results. Because of the special structure of the transportation graph, namely the nodes can be separated into two classes (sources and sinks) so that all links join nodes of the first class to nodes of the second class, the 0-1 incidence matrix  $H$  has a very special property, as we shall see later on.

All we will need concerning linear graphs are the following notions. For the moment, denote the nodes of a graph  $G$  by  $1, 2, \dots, N$  and the links by unordered pairs of nodes. Then a chain is a set of one or more links that can be arranged as  $i_1 i_2, i_2 i_3, \dots, i_{k-1} i_k$  where the nodes  $i_1, i_2, \dots, i_k$  are distinct. A cycle is a set of links that can be ordered as  $i_1 i_2, i_2 i_3, \dots, i_{k-1} i_k, i_k i_1$ , the nodes  $i_1, i_2, \dots, i_k$  again being distinct. A graph is connected if each pair of nodes is joined by a chain, and a connected graph without cycles is a

tree. Equivalently, a tree is a graph in which each pair of nodes is joined by just one chain, since the union of two chains joining the same pair of nodes contains a cycle. Let  $G$  have  $L$  links and  $P$  connected pieces; then  $L - N + P$ , known as the cyclic number of  $G$ , gives the maximal number of independent cycles in  $G$ , where a set of cycles is independent if each contains a link not in any of the others. Thus any two of the three conditions (a)  $G$  connected, (b)  $G$  has no cycles, (c)  $L = N - 1$ , imply the third and characterize  $G$  as a tree.

Now corresponding to any basis for the transportation problem, we may associate a subgraph of the full transportation graph by deleting those links corresponding to non-basic variables. It is now easy to prove the following fundamental theorem, which accounts for the simplicity of the simplex computation for transportation problems.

Theorem. A set of variables is a basic set if and only if its graph is a tree.

Proof. If  $B$  is a basic set, Lemma 2 shows that  $B$  contains  $m + n - 1$  variables. Since the graph  $G(B)$  associated with  $B$  has  $m + n$  nodes, we see that condition (c) above is fulfilled. We now verify (b). Notice that any cycle of the transportation graph must be of the form

$$i_1 j_1, j_1 i_2, \dots, i_k j_k, j_k i_1.$$

Thus, since

$$x_{i_1 j_1} - x_{i_2 j_1} + \dots + x_{i_k j_k} - x_{i_1 j_k} = 0.$$

$G(B)$  can contain no cycles. It is therefore a tree.

Conversely, let  $B$  be any set of variables whose graph  $G(B)$  is a tree. By (c),  $B$  contains the proper number,  $m + n - 1$ , of variables. Moreover, the associated column vectors of  $H$  are linearly independent. To see this, locate an "end-node" in  $G(B)$  (i.e. a node incident with just one link of  $G(B)$ ); there must be at least two such. Let this node be a source-node  $\bar{i}$ , say, and suppose  $\bar{i}\bar{j}$  is the link. Then, assuming a linear dependence

$$\sum_{1j \in G(B)} k_{1j} x_{1j} = 0,$$

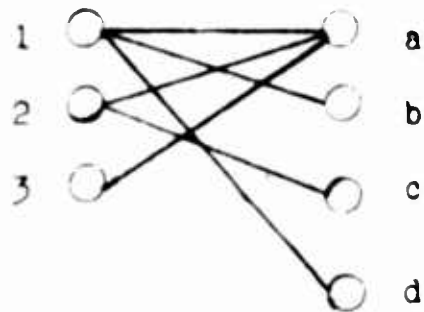
we deduce immediately that  $k_{\bar{i}\bar{j}} = 0$ . We then delete this node and the link on it, leaving a tree, and repeat the procedure. It follows that all the coefficients of the assumed linear dependence vanish, and hence  $B$  is a basis, although of course not necessarily a feasible one. (Notice that Lemma 2 was proved by selecting a particular subtree.)

We point out that this last part of the argument does not rely on any special features of the transportation graph. In other words, the incidence matrix of nodes vs. links of any tree has maximal rank. Moreover, the argument also shows that such a matrix may be arranged in triangular form, and indeed gives a prescription for doing so.

Corollary 1. Every transportation basis is triangular.



Example.



	(1d)	(2c)	(1b)	(3a)	(2a)	(1a)
d	1	0	0	0	0	0
c	0	1	0	0	0	0
b	0	0	1	0	0	0
3	0	0	0	1	0	0
2	0	1	0	0	1	0
a	0	0	0	1	1	1
1	1	0	1	0	0	1

Corollary 2. If the  $a_1$  and  $b_j$  are integers, so are the  $x_{ij}$  for any basic solution.

Proof. Since the coefficients in the equations are either 0 or 1, and every basis is triangular, only addition and subtraction are required in solving for a basic solution.

Actually one can say somewhat more — the value of any variable in a basic solution is either a partial sum of the  $a_1$  minus a partial sum of the  $b_j$ , or vice versa. Thus

Corollary 3. Degeneracy in the simplex computation can occur only if a partial sum of the  $a_{ij}$  equals a partial sum of the  $b_j$ .

We next turn attention to the simplex multipliers. Denote by  $\alpha_i$  ( $\beta_j$ ) the multiplier corresponding to the  $i^{\text{th}}$  source ( $j^{\text{th}}$  sink) equation. Then the equations to be solved in evaluating the simplex multipliers are

$$(3) \quad \alpha_i + \beta_j = a_{ij} \quad (ij \in G(B)).$$

Thus we may think of assigning a number to each node so that the sum of any two numbers on nodes joined by a link of  $G(B)$  is equal to the transportation cost for the link. Because one of the original equations is redundant, (3) determines the  $\alpha_i$  and  $\beta_j$  only to within an additive constant; more precisely, if  $\alpha_i, \beta_j$  satisfy (3), all solutions are given by  $\alpha_i - k, \beta_j + k$ . Fixing some multiplier at zero, say, gives a unique determination. From Corollary 1 we have

Corollary 4. If the  $a_{ij}$  are integers and one of the multipliers  $\alpha_i, \beta_j$  is assigned an integral value, all multipliers are integral.

Let source  $i$  and sink  $j$  be joined by the (unique) chain, necessarily of odd length,

$$i j_1, j_1 i_1, i_1 j_2, \dots, j_k i_k, i_k j$$

in  $G(B)$ . Then, solving for the multipliers along this chain,

we see that

$$\beta_{j_1} = a_{1j_1} - \alpha_1$$

$$\alpha_{1_1} = a_{1_1j_1} - a_{1j_1} + \alpha_1$$

$$\beta_{j_2} = a_{1_1j_2} - a_{1_1j_1} + a_{1j_1} - \alpha_1$$

$$\dots \dots \dots \beta_j = a_{1_kj} - a_{1_kj_k} + \dots + a_{1_1j_2} - a_{1_1j_1} + a_{1j_1} - \alpha_1.$$

Hence

Corollary 5. The simplex multipliers satisfy

$$\alpha_1 + \beta_j = a_{1j_1} - a_{1_1j_1} + a_{1_1j_2} - \dots - a_{1_kj_k} + a_{1_kj}$$

where  $1j_1, j_11_1, 1_1j_2, \dots, j_k1_k, 1_kj$  is the chain joining 1 and j in  $G(B)$ .

The simplex criterion for optimality of a basic feasible solution is

$$(4) \quad \delta_{1j} = a_{1j} - \alpha_1 - \beta_j \geq 0$$

for all non-basic variables  $x_{1j}$ . If some  $\delta_{1j} < 0$ , the variable  $x_{1j}$  may be introduced into the basic set with a possible decrease in cost. To introduce such a variable, we need only look for the chain joining 1 and j; this chain, together with link  $1j$ , constitutes a cycle; it is the only cycle in the graph  $G(B) \cup (1j)$ . For a cycle, as we have seen, the associated column vectors of  $H$  are linearly dependent, the coefficients of the dependence being alternately +1 and -1. Stated another way, if the cycle is

$$j_1, 1j_1, j_1^1, \dots, j_k^1, 1_kj$$

and we denote the basic feasible solution by  $\hat{x}$ , then a change of  $\theta$  in  $\hat{x}_{1j}$  is compensated for by alternately subtracting and adding  $\theta$  along the chain, i.e.

$$(5) \quad x_{1j} = \theta, x_{1j_1} = \hat{x}_{1j_1} - \theta, x_{1_1j_1} = \hat{x}_{1_1j_1} + \theta, \dots, x_{1_kj} = \hat{x}_{1_kj} - \theta.$$

Using Corollary 5, it is clear that if

$$\delta_{1j} = a_{1j} - a_{1j_1} + a_{1_1j_1} - \dots + a_{1_kj_k} - a_{1_kj} < 0,$$

this change results in a decrease of  $\theta\delta_{1j}$  in the transportation cost.

The new set of basic variables results by taking  $\theta$  as large as possible consistent with maintaining non-negativity of variables, i.e.

$$(6) \quad \theta = \min ( \hat{x}_{1j_1}, \hat{x}_{1_1j_2}, \dots, \hat{x}_{1_kj} ) .$$

Now it may happen that the minimum in (6) occurs for more than one index pair. This situation, referred to as a "tie" in the algorithm, might lead, on some future basis change, to a determination of  $\theta = 0$ , i.e. to degeneracy in the algorithm, and the consequent possibility of non-termination. By Corollary 3, this can happen only if a partial sum of  $a_i$  equals a partial sum of  $b_j$ . Viewing Corollary 3 in a slightly different way, we see that if a tie has occurred, the graph of the strictly positive basic variables in the new set is disconnected, and thus the

original problem splits up into feasible "sub-problems."

It is easy to see from Corollary 3 that degeneracy can be avoided, if one desires, by perturbing the  $a_i$  and  $b_j$  as follows:

$$a_1 \longrightarrow a'_1 = a_1 + n\varepsilon$$

$$a_i \longrightarrow a'_i = a_i \quad (i > 1)$$

$$b_j \longrightarrow b'_j = b_j + \varepsilon.$$

For  $\varepsilon$  sufficiently small, no partial sum of the  $a'_i$  can equal a partial sum of the  $b'_j$ , as otherwise we would have either  $0\varepsilon = k\varepsilon$  or  $n\varepsilon = k\varepsilon$ , where  $1 \leq k \leq n-1$ , a contradiction. A value of  $\varepsilon$  need never be specified; when a tie occurs in the original problem, it can be resolved by evaluating the " $\varepsilon$ -part" of the solution. Another way to resolve ties is just to select at random among the tied alternatives. It is easy to see that the simplex computation then terminates with probability one.

It may be instructive at this point to re-interpret the various steps of the simplex computation in terms of the array (instead of a linear graph) commonly used for hand computation:

		Sinks								
		1	2	3	4	5	6	7		
Sources	1	$x_{11} - \theta$ $a_{11}$	$x_{12} + \theta$ $a_{12}$	$a_{13}$	$a_{14}$	$a_{15}$	$a_{16}$	$a_{17}$	$a_1$ $\alpha_1$	1
	2	$a_{21}$	$x_{22} - \theta$ $a_{22}$	$a_{23}$	$x_{24} + \theta$ $a_{24}$	$a_{25}$	$x_{26}$ $a_{26}$	$a_{27}$	$a_2$ $\alpha_2$	2
	3	$\theta$ $a_{31}$	$a_{32}$	$a_{33}$	$a_{34}$	$x_{35} - \theta$ $a_{35}$	$a_{36}$	$a_{37}$	$a_3$ $\alpha_3$	5
	4	$a_{41}$	$a_{42}$	$x_{43}$ $a_{43}$	$x_{44} - \theta$ $a_{44}$	$x_{45} + \theta$ $a_{45}$	$a_{46}$	$x_{47}$ $a_{47}$	$a_4$ $\alpha_4$	4
	5	$b_1$ $\beta_1$	$b_2$ $\beta_2$	$b_3$ $\beta_3$	$b_4$ $\beta_4$	$b_5$ $\beta_5$	$b_6$ $\beta_6$	$b_7$ $\beta_7$		
		0	1	4	2	4	2	4	←	Labeling ↑

The values of the current basic set are recorded in their appropriate cells (indicated in the example by the  $x_{ij}$  shown). To compute the multipliers  $\alpha_i, \beta_j$ , assign  $\alpha_1 = 0$ , say, then scan the first row for basic cells, computing the corresponding  $\beta_j$  ( $\beta_1$  and  $\beta_2$  in the example). Then scan the columns associated with newly computed  $\beta_j$ , evaluating a set of  $\alpha$ 's ( $\alpha_2$  in the example). Continue this process until all multipliers are evaluated. The next step is to compute all  $b_{ij} - a_{ij} - \alpha_i - \beta_j$  and find the most negative one (assumed to be  $b_{31}$  in the example). Increasing this variable to  $\theta$  causes the chain of reactions indicated by  $+\theta$  and  $-\theta$  in the array. Taking  $\theta$  as large as possible ( $\min(x_{11}, x_{22}, x_{44}, x_{35})$ ) gives the new basic feasible solution. Notice that the cycle one is looking for takes the form of a "rook's tour" in the array. For small problems, searching for a cycle is

easily done without any special algorithm. For larger problems, however, one needs an efficient algorithm for this search. We shall describe one such, which might be referred to as the "labeling procedure" [8]. This procedure is, in general, an efficient way of finding a chain joining two given nodes in an arbitrary network, and will play an important role in subsequent lectures on network problems and a new algorithm for the transportation problem.

Label the column which contains the variable to be introduced by a 0, say, and scan the column for basic cells; when one is found, label the corresponding row by the number of the column being scanned; take any newly labeled row and scan it for basic cells contained in unlabeled columns; label the corresponding columns by the number of the row being scanned; when these rows have all been scanned, go to the newly labeled columns and repeat the procedure (never labeling a row or column which already has a label), stopping only when the row which contains the cell being introduced has been labeled. Then proceed, in that row, to the position indicated by its label, and record  $-\theta$ ; then go, in the column reached, to the position indicated by its label, and record  $+\theta$ . Continue this backward replacement until the original column, labeled 0, is reached.

A labeling which finds the desired cycle is indicated at the bottom and to the right of the array in the example.

For machine calculation, it is probably a good idea to carry along a separate list of basic cells, sorted perhaps on both rows and columns, and to do the computation of the multi-

pliers and the search for the cycle in terms of just this list.

So far we have said nothing about finding a starting basic feasible solution. This again is an easy task. Simply select any cell  $ij$  and set

$$x_{ij} = \min (a_i, b_j).$$

If  $x_{ij} = a_i < b_j$ , delete the  $i^{\text{th}}$  row, replace  $b_j$  by  $b_j - a_i$ , and select another cell in the reduced array. If  $x_{ij} = b_j < a_i$ , delete the  $j^{\text{th}}$  column and reduce  $a_i$  by  $b_j$  to obtain the reduced array. If  $a_i = b_j$ , delete either the row or column, unless there are several rows (columns) left and only one column (row), in which case a row (column) is deleted.

This completes the discussion of the simplex computation for transportation problems.

## 2. OPTIMAL ASSIGNMENT PROBLEM

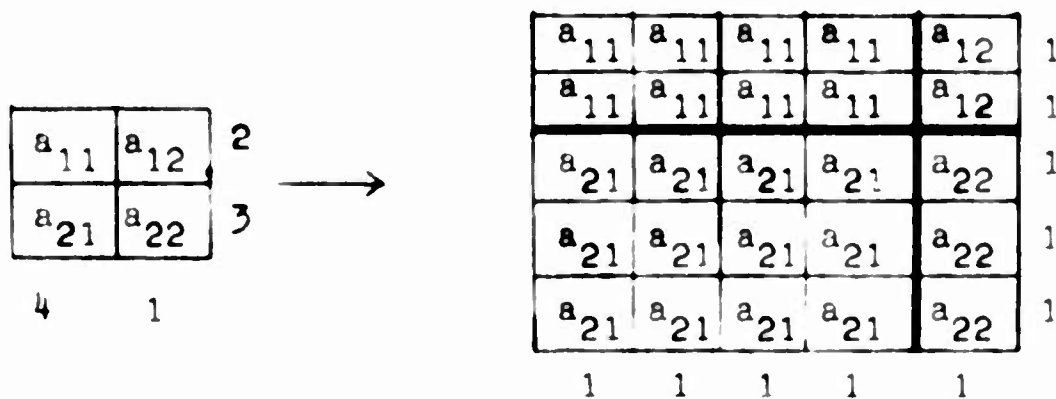
As we have mentioned earlier, the integrality of basic feasible solutions for the Hitchcock problem with integral supplies and demands enables one to solve some combinatorial problems by setting them up as Hitchcock problems. Perhaps the best known example of this is the problem of optimal assignment (1a in the list), where the constraints take the form

$$(7) \quad \begin{aligned} \sum_j x_{ij} &= 1 & (i = 1, \dots, M) \\ \sum_i x_{ij} &= 1 & (j = 1, \dots, M) \\ x_{ij} &\geq 0 \end{aligned}$$



and we wish to maximize  $\sum_{i,j} a_{ij}x_{ij}$ . Since basic feasible solutions are integral, each row and each column of such a solution contains a 1 and 0's elsewhere, and this describes a permutation or assignment of men to jobs. Conversely, any permutation matrix is a basic feasible solution. Thus an optimal assignment problem may be solved as a Hitchcock problem having the constraints (7).

It is also true that a Hitchcock problem with integral  $a_i$  and  $b_j$  may be formulated as an optimal assignment problem with  $M = \sum a_i$ . The idea can be made clear by considering an example (pictured schematically below):



Thus the  $i^{th}$  source is replaced by  $a_i$  sources, each with unit supply and the same shipping costs, and the process is repeated for the sinks.

### 3. UPPER BOUNDS ON VARIABLES

In addition to the constraints (2), one may add "capacity" restrictions

$$(8) \quad x_{ij} \leq c_{ij}$$

without altering the character of the problem, except that the

feasibility criterion is no longer simple. Since upper bounds on variables will be dealt with later on in the course for the general linear programming problem, we will not discuss the matter here in any detail. Briefly, the situation is this. If inequalities (8) are rewritten in the usual way as

$$(9) \quad x_{1j} + y_{1j} = c_{1j}, \quad y_{1j} \geq 0,$$

then one of  $x_{1j}$ ,  $y_{1j}$  must be in any basic set. Thus, if  $B$  denotes a basic set, the pairs  $1j$  fall into three classes: those  $1j$  for which  $x_{1j} \in B$ ,  $y_{1j} \in B$ ; those for which  $x_{1j} \in B$ ,  $y_{1j} \notin B$  (hence  $x_{1j} = c_{1j}$ ); and those for which  $x_{1j} \notin B$ ,  $y_{1j} \in B$  (hence  $y_{1j} = c_{1j}$ ). The graph of the pairs of the first kind can be shown to be a tree, and thus the basis is again triangular.

#### 4. THE TRANSSHIPMENT PROBLEM

The transshipment problem (problem 2 in the list) arises frequently in practical applications. It was originally formulated and solved by A. Orden [16]; his method of solution is to add the "transshipment variables"  $x_{11}$  and double the original number of equations, thereby getting the problem in the standard Hitchcock form. Instead of using this device, we shall give a brief discussion of the problem directly in terms of the given linear graph.

If we let  $x_{1j}$  ( $1, j = 1, \dots, N$ ) denote the commodity flow from node 1 to node  $j$  and assume that the numbering of the nodes is such that  $1, \dots, k$  are sources,  $k + 1, \dots, N$  are sinks,

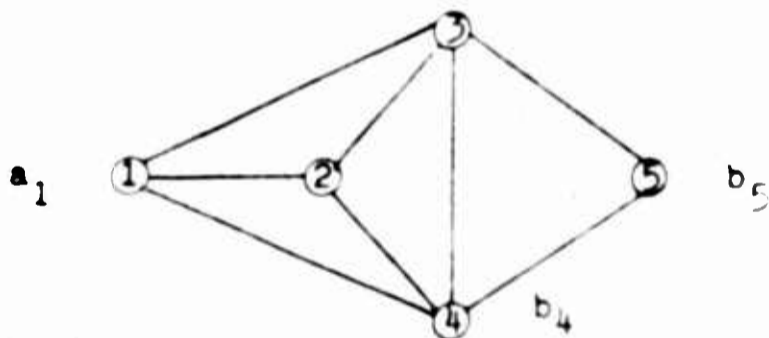
then the constraints are

$$\begin{aligned}
 (10) \quad \sum_{j \neq 1} (x_{j1} - x_{1j}) &= -a_1 & (i = 1, \dots, k) \\
 \sum_{j \neq 1} (x_{j1} - x_{1j}) &= 0 & (i = k + 1, \dots, k + l - 1) \\
 \sum_{j \neq 1} (x_{j1} - x_{1j}) &= b_1 & (i = k + l, \dots, N) \\
 x_{ij} &\geq 0
 \end{aligned}$$

(where variables corresponding to missing links do not appear),  
 subject to which it is desired to minimize

$$(11) \quad \sum_{\substack{i,j \\ i \neq j}} a_{ij} x_{ij}$$

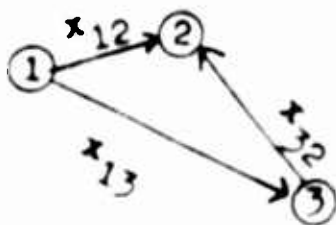
Thus, for example, if the network is



the matrix T of the linear programming problem is

	$x_{12}$	$x_{21}$	$x_{13}$	$x_{31}$	$x_{14}$	$x_{41}$	$x_{23}$	$x_{32}$	$x_{24}$	$x_{42}$	$x_{34}$	$x_{43}$	$x_{35}$	$x_{53}$	$x_{45}$	$x_{54}$
1	-1	1	-1	1	-1	1	0	0	0	0	0	0	0	0	0	0
2	1	-1	0	0	0	0	-1	1	-1	1	0	0	0	0	0	0
3	0	0	1	-1	0	0	1	-1	0	0	-1	1	-1	1	0	0
4	0	0	0	0	1	-1	0	0	1	-1	1	-1	0	0	-1	1
5	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	1	-1

Notice that the column vector  $X_{ij}$  has  $-1$  in the  $i^{\text{th}}$  position,  $+1$  in the  $j^{\text{th}}$  position, and zeros elsewhere. Consequently the sum of the rows of  $T$  vanishes, and  $\text{rank } T \leq N - 1$ . Just as in the Hitchcock problem, we can exhibit a set of  $N - 1$  linearly independent column vectors of  $T$  (provided the graph is connected, as we assume to insure feasibility) by selecting a spanning subtree of the graph and choosing either  $X_{ij}$  or  $X_{ji}$  corresponding to the links of the tree. Indeed, every set of  $N - 1$  linearly independent column vectors can be exhibited in this fashion, i.e. the fundamental theorem on the correspondence of basic sets of variables and trees remains valid.\* The proof is little different; one just observes that because  $X_{ij}$  has  $-1$  in the  $i^{\text{th}}$  component,  $+1$  in the  $j^{\text{th}}$ , any choice of columns corresponding to a cycle leads to a dependency. More specifically, if we orient the links of the cycle  $C$  in the directions dictated by the choice of variables, e.g.



then  $\sum_{ij \in C} \pm X_{ij} = 0$ , where the  $-$  sign is selected if, in

traversing the cycle, link  $ij$  is traversed in the direction

---

\*More properly, a basic set of variables corresponds to an oriented spanning subtree of the graph, since the incidence matrix  $T$  is composed of ones, minus ones, and zeros.

P-890  
 7-9-56  
 -24-

opposite its orientation, say, the + sign otherwise.

One difference between the transshipment and Hitchcock problems that should be pointed out is that the convex polyhedron defined by (10) is unbounded. The assumption  $a_{ij} \geq 0$  guarantees a finite minimum for the problem, however.

The simplex computation for the transshipment problem differs only in detail from the computation for the Hitchcock problem. The simplex multipliers, one for each node, are found by assigning one of them an arbitrary value, and computing the others along the tree, taking account of the orientation of the links. Introducing a variable  $x_{ij}$  into a basic set is again equivalent to finding the chain of the tree which joins node  $i$  to node  $j$ , and may be done by a labeling procedure.

For machine computation, one would of course not have a separate program for the transshipment problem, but would use Orden's approach and solve an equivalent  $N$  by  $N$  Hitchcock problem. For the transshipment graph pictured earlier, the equivalent Hitchcock problem is described by the array

	1	2	3	4	5	
1	0	$a_{12}$	$a_{13}$	$a_{14}$	$\infty$	$s + a_1$
2	$a_{21}$	0	$a_{23}$	$a_{24}$	$\infty$	$s + a_2$
3	$a_{31}$	$a_{32}$	0	$a_{34}$	$a_{35}$	$s$
4	$a_{41}$	$a_{42}$	$a_{43}$	0	$a_{45}$	$s$
5	$\infty$	$\infty$	$a_{53}$	$a_{54}$	0	$s$
	$s$	$s$	$s$	$s + b_4$	$s + b_5$	

P-890  
7-9-56  
~~-25-~~

where cells with costs  $+\infty$  correspond to missing variables,  
 $x_{11}$  has zero cost, and  $s$  is a sufficiently large number ( $s = \sum a_1$   
will do).

REFERENCES

1. Cahn, A. S., "The warehouse problem," Bull. Amer. Math. Soc., vol. 54, Oct. 1948.
2. Dantzig, G. B., "Application of the simplex method to a transportation problem," Activity Analysis of Production and Allocation, Cowles Commission Monograph No. 13, T. C. Koopmans, Ed., John Wiley and Sons, Inc., New York, 1951.
3. Dantzig, G. B., and D. R. Fulkerson, "Minimizing the number of tankers to meet a fixed schedule," Naval Research Logistics Quarterly, vol. 1, no. 3, Sept. 1954.
4. \_\_\_\_\_, "On the min cut max flow theorem of networks," RAND research memorandum RM-1418, Jan. 1955 (to appear in Contributions to Linear Inequalities and Related Topics, Annals of Math. Study No. 38).
5. \_\_\_\_\_, "Computation of maximal flows in networks," Naval Research Logistics Quarterly, vol. 2, no. 4, Dec. 1955.
6. Dreyfus, S. and T. F. Cartaino, "Application of dynamic programming to the airplane minimum time to climb problem," RAND paper P-834, March, 1956.
7. Ford, L. R. Jr. and D. R. Fulkerson, "Maximal flow through a network," RAND paper P-605, Nov. 1954 (to appear in Canadian Journal of Math.)
8. \_\_\_\_\_, "A simple algorithm for finding maximal network flows and an application to the Hitchcock problem," RAND paper P-743, Sept. 1955.
9. Fulkerson, D. R., I. Glicksberg, and O. Gross, "A production line assignment problem," RAND research memorandum RM-1102, May 1953.
10. Gale, D., "A theorem on flows in networks" (to appear in Pacific Journal Math.)
11. Hall, P. "On representatives of subsets," Jour. London Math. Soc., vol. 10, 1935.
12. Hitchcock, F. L., "The distribution of a product from several sources to numerous localities," Jour. Math. Phys., vol. 20, 1941.
13. Hoffman, A. and G. B. Dantzig, "Dilworth's theorem on partially ordered sets," RAND research memorandum RM-1553, Aug. 1955

(to appear in Contributions to Linear Inequalities and related topics, Annals of Math. Study No. 38).

14. Jacobs, W., "The caterer problem," Naval Research Logistics Quarterly, vol. 1, 1954.
15. Koopmans, T. C. and S. Reiter, "A model of transportation," Activity Analysis of Production and Allocation, Cowles Commission monograph No. 13, T. C. Koopmans, Ed., John Wiley and Sons, Inc. New York, 1951.
16. Orden, A., "The transshipment problem," Management Science, vol. 2, no. 3, April 1956.
17. Prager, W., "On the caterer problem," Brown Univ., April, 1956.
18. Smith, L. W. Jr., "Current status of the industrial use of linear programming," Management Science, vol. 2, no. 2, Jan. 1956.
19. Votaw, D. F. Jr., and A. Orden, "The personnel assignment problem," Project Scoop symposium on Linear Inequalities and Programming, Washington, D. C., April 1952.