

SOLUTION OF A LARGE SCALE TRAVELING  
SALESMAN PROBLEM

by  
G. Dantzig, R. Fulkerson  
and  
S. Johnson

P-510

12 April 1954

---

*The* **RAND** *Corporation*

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

---

#### SUMMARY

It is shown that a certain tour of 49 cities, one in each of the 48 states and Washington, D. C. has the shortest road distance.

### Historical Note

The origin of this problem is somewhat obscure. It appears to have been discussed informally among mathematicians at mathematics meetings for many years. Surprisingly little in the way of results has appeared in the mathematical literature [9]. It is likely that the minimal distance tour problem was stimulated by the so-called Hamiltonian Game [7] which is concerned with finding the number of different tours possible over a specified network. The latter problem is credited by some as the origin of group theory and has some connections with the famous Four Color Conjecture [8]. Merrill Flood (Columbia University) should certainly be credited with stimulating interest in this traveling salesman problem in many quarters. As early as 1937, he tried to obtain near optimal solutions in reference to routing of school buses. Both Flood and A. W. Tucker (Princeton University) recall that they heard about the problem first in a seminar talk by Hassler Whitney at Princeton in 1934 (although Whitney, recently queried, does not seem to recall the problem). The relations between the traveling salesman problem and the transportation problem of linear programming appear to have been first explored by M. Flood, J. Robinson, T. C. Koopmans, and later by H. Kuhn and I. Heller [3,4].

SOLUTION OF A LARGE SCALE TRAVELING  
SALESMAN PROBLEM

by

G. Dantzig, R. Fulkerson  
and  
S. Johnson

1. Introduction. The traveling salesman problem might be described as follows: Find the shortest route for a salesman starting from a given city, visiting each of a specified group of cities, and then returning to the original point of departure. More generally, given an  $n$  by  $n$  symmetric matrix  $D = (d_{ij})$ , where  $d_{ij}$  represents the "distance" from  $i$  to  $j$ , arrange the points in a cyclic order in such a way that the sum of the  $d_{ij}$  between consecutive points is minimal. Since there are only a finite number of possibilities (at most  $\frac{(n-1)!}{2}$ ) to consider, the problem is to devise a method of picking out the optimal arrangement which is reasonably efficient for fairly large values of  $n$ . Although algorithms have been devised for problems of similar nature, e.g., the optimal assignment problem [2,5,6], little is known about the traveling salesman problem. We do not claim that this note alters the situation very much; what we shall do is outline a way of approaching the problem that sometimes, at least, enables one to find an optimal path and prove it so. In particular, it will be shown that a certain arrangement of 49 cities, one in each of the 48 states and Washington, D. C. is best, the  $d_{ij}$  used representing road distances as taken from a Rand McNally atlas.

We attack the problem essentially by using a graphical interpretation of the simplex algorithm, that is, we formulate a related linear programming problem, or more precisely, a sequence of programming problems. The linear programming approach to combinatorial problems is of course nothing new; its most marked success has perhaps been in the optimal assignment problem. There the problem is, given an  $n$  by  $n$  matrix  $a_{ij}$ , find a permutation  $p^*$  which achieves  $\max_p \sum_{i=1}^n a_{i,p(i)}$ . The underlying fact is that the permutation matrices are the only extreme points of the convex set (in  $n^2$ -space) of doubly stochastic matrices, and consequently one can state the problem in the programming form: Find  $\max_{x_{ij}} \sum_{i,j} a_{ij} x_{ij}$ , where  $\sum_{i=1}^n x_{ij} = \sum_{j=1}^n x_{ij} = 1$ ,  $x_{ij} \geq 0$ . The difficulty in using an analogous procedure for the traveling salesman problem has been emphasized by work of H. Kuhn [4] and I. Heller [3]. The extreme hyperplanes of the convex spanned by those permutations which are  $n$ -cycles (called tours from now on) are not only hard to find, even for small  $n$ , but are so profuse in number that a straight forward programming approach would not be feasible, even if one knew them all. We try to get around this difficulty in two ways. First of all, by utilizing the fact that the distance matrix is symmetric, one can map the  $n^2$ -space onto a lower dimensional space in such a way that tours which differ only in direction of traversal are identified; this seems to shorten the computation and also appears to introduce some simplification, at least for small  $n$ , into a characterization of the convex  $T_n$  of tours. Secondly, and

most important, we try to find a convex  $C \supset T_n$  which has the tours as extreme points (and many others), but over which the minimum of the linear form is assumed at a tour. For the particular 49-city problem mentioned, and also for all the smaller problems we have considered, such a convex  $C$  has been relatively easy to find by hand computation.

2. Preliminary notions. A (directed) tour for  $n$  points can be thought of as a permutation matrix of order  $n$  which represents an  $n$ -cycle. For example, for  $n=5$ , the matrix (1) below

$$(1) \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} ; \quad (2) \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

is a tour, since it represents the 5-cycle (12435), while the matrix (2) above is not a tour since it represents the permutation (12)(354). Now define a linear mapping from the  $n^2$ -space of  $n$  by  $n$  matrices  $x' = (x'_{ij})$  onto the  $\frac{n(n-1)}{2}$ -space of triangular arrays  $x = (x_{ij})$ ,  $i > j$ , by  $L(x') = x$  where  $x_{ij} = x'_{ij} + x'_{ji}$ . Thus the images under  $L$  of the two displayed matrices would be the arrays (or vectors)

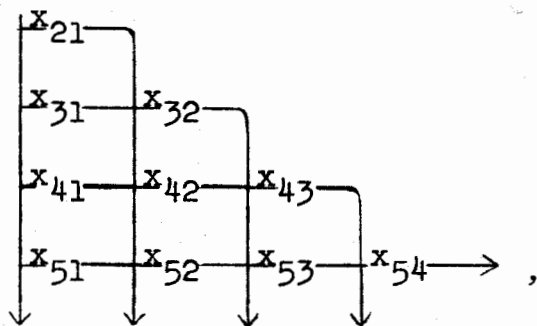
$$(3) \quad \begin{bmatrix} 1 & & & \\ 0 & 0 & & \\ 0 & 1 & 1 & \\ 1 & 0 & 1 & 0 \end{bmatrix}, \quad (4) \quad \begin{bmatrix} 2 & & & \\ 0 & 0 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

respectively. Any image under  $L$  of a (directed) tour in the  $n^2$ -space will be called an undirected tour in the  $\frac{n(n-1)}{2}$ -space, or briefly, a tour. This definition is meaningful since the two (directed) tours  $(i_1, i_2, \dots, i_n)$  and  $(i_n, i_{n-1}, \dots, i_1)$  clearly have the same image under  $L$ . The reason we can consider tours in the lower dimensional space is of course that we assume the form to be minimized is symmetric, hence all information is embodied in a triangular array of distances.

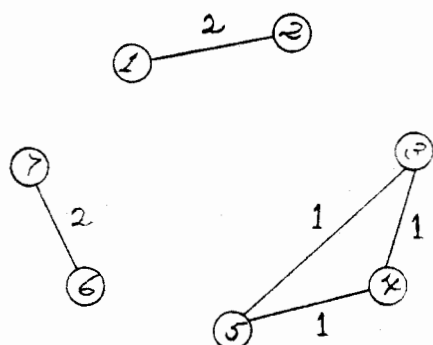
The image under  $L$  of the convex polyhedron  $C'_1$  given by  $\sum_i x'_{ij} = \sum_j x'_{ij} = 1$ ,  $x'_{ii} = 0$ ,  $x'_{ij} \geq 0$ , is the convex polyhedron  $C_1$  defined by

$$(2.1) \quad \left\{ \begin{array}{l} \sum_{i \text{ or } j=1_0} x_{ij} = 2, \quad i_0 = 1, \dots, n, \\ x_{ij} \geq 0. \end{array} \right. \quad (i > j)$$

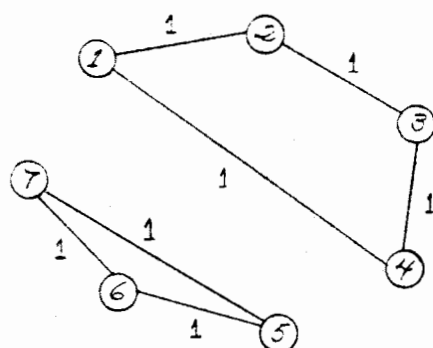
This will essentially be the starting polyhedron in the problems to be discussed later. The equations in (2.1) can be pictured as



or, in terms of the map of points and the weighted links joining them, as saying that the sum of the weights on the segments emanating from each point is 2. Now the extreme points of  $C_1'$  are just the permutation matrices which leave no integer fixed. Consequently, since extreme points of  $C_1$  must be images under  $L$  of extreme points of  $C_1'$ , the only possibilities for extreme points of  $C_1$  are triangular arrays consisting of 0's, 1's, and 2's, that is to say, maps (where line segments having 0-weight are omitted) which consist of line segments with weight 2 and "loops" or sub-tours having weight 1 on each of the segments comprising them. For example, for  $n = 7$ , the point  $x$  represented by the graph



is a candidate for an extreme point of  $C_1$ , and is, in fact, an extreme point. On the other hand, the point  $y$  represented by





is not an extreme point, the reason being that it contains a loop with an even number of sides. In other words, a point is extreme in  $C_1$  if and only if its graph has no link with non-integral weight and contains only odd loops. The rest of the proof of this statement can be made clear by considering the examples. Suppose  $x = 1/2(u+v)$ . Then  $u$  and  $v$  must have 0's and 2's where  $x$  does. But

$$u_{43} = 2 - u_{53}, u_{54} = u_{53}, \text{ and so } u_{53} + u_{54} = 2u_{54} = 2, u_{54} = 1.$$

Consequently  $x = u = v$ . On the other hand, the point  $y$  for the second example is the midpoint of the line segment joining  $u, v \in C_1$ , where  $u_{21} = 1 \frac{1}{2}, u_{32} = 1/2, u_{43} = 1 \frac{1}{2}, u_{41} = 1/2, u_{75} = u_{65} = u_{76} = 1$ , and  $v_{21} = 1/2, v_{32} = 1 \frac{1}{2}, v_{43} = 1/2, v_{41} = 1 \frac{1}{2}, v_{75} = v_{65} = v_{76} = 1$ .

Thus one result of reducing the dimension of the space is that tours are extreme points of  $C_1$  for odd  $n$  but are not extreme points of  $C_1$  for even  $n$ . An even number of points is really no obstacle, however, since if the elements of the array  $D = (d_{ij})$ ,  $i > j$ , satisfy the triangle inequality, simply count one of the points twice. It is not difficult to show that the triangle inequality for triangles having this point as a vertex implies that some optimal tour goes from this point to "itself." Another way of avoiding the difficulty is to impose upper bounds on the variables,  $x_{ij} \leq 1$ . It is easy to see that the new convex defined admits all tours as extreme points.

Another obvious set of relations that tours satisfy, in addition to those already mentioned, is the following: Partition the set  $I = \{1, 2, \dots, n\}$  of points into two complementary subsets  $I_1$  and  $I_2$

where each contains at least two elements. Let  $\Sigma_{1,2}$  denote the sum of the weights of all line segments joining points of  $I_1$  with points of  $I_2$ . For any tour,  $\Sigma_{1,2}$  must be an even integer, hence all tours lie in the half-spaces

$$(2.2) \quad \Sigma_{1,2} \geq 2 \quad .$$

The only reason for supposing  $I_1$  and  $I_2$  contain two or more points is that if  $I_1$ , say, is a point, we already have the conditions ( recall (2.1) ),  $\Sigma_{1,2} = 2$ . Now for any such partitioning, let  $\Sigma_1$  denote the sum of the weights of all links connecting the  $n_1$  points of  $I_1$ . Define  $\Sigma_2$  similarly. Then (2.2) is equivalent to either of the conditions

$$(2.3) \quad \Sigma_1 \leq n_1 - 1 \quad ,$$

$$(2.4) \quad \Sigma_2 \leq n_2 - 1 \quad ,$$

by virtue of (2.1). This follows by observing, for example, that  $2 \Sigma_1 + \Sigma_{1,2} = 2n_1$ , since if one adds up the equations of (2.1) which refer to points of  $I_1$ , the segments in  $\Sigma_1$  are counted twice, those in  $\Sigma_{1,2}$  only once. Thus the set of all inequalities of type (2.2) is equivalent to the set of all inequalities of type (2.3), where  $I_1$  runs over all subsets of  $n_1$  points with  $1 < n_1 \leq \frac{n}{2}$ . Note that upper bounds  $x_{1j} \leq 1$  are a special case of (2.2).

It turns out, for  $n \leq 5$ , that relations of type (2.1) and (2.2) suffice to characterize  $T_n$ . Thus  $T_5$  can be described as the convex in 10-space defined by the 5 equations of (2.1) and the 20 inequalities  $0 \leq x'_{ij} \leq 1$ , whereas  $T'_5$ , the convex of directed tours, has the irredundant characterization [3],

$$(a) \sum_i x'_{ij} = \sum_j x'_{ij} = 1 \text{ (one equation omitted to avoid redundancy)}$$

$$(b) x'_{ij} \geq 0$$

$$(c) x'_{ii} = 0$$

$$(d) x'_{ij} + x'_{ji} \leq 1$$

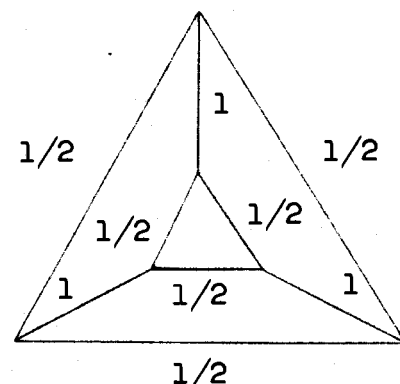
$$(e) x'_{ij} + x'_{ji} + x'_{rs} - x'_{st} - x'_{tr} \leq 1 \text{ for distinct } (i, j, r, s, t)$$

$$(f) 2x'_{ij} + 2x'_{ji} - x'_{ir} + x'_{jr} - x'_{si} + x'_{sj} - x'_{rs} \leq 2 \text{ for distinct } (i, j, r, s),$$

a system of 224 hyperplanes in 25-space. This makes it appear that the situation has been simplified somewhat by the mapping  $L$ , at least for  $n \leq 5$ . We have not attempted any detailed analysis of  $T_n$  for  $n > 5$  since we have seldom, in working problems, been forced to go outside relations of types (2.1) and (2.2). It is true, of course, that these relations do not define  $T_n$  for  $n > 5$ . For example, for  $n = 6$ , the "fractional" point

$$\begin{bmatrix} 1/2 & & & & \\ 1/2 & 1/2 & & & \\ 1 & 0 & 0 & & \\ 0 & 1 & 0 & 1/2 & \\ 0 & 0 & 1 & 1/2 & 1/2 \end{bmatrix}$$

or



is an extreme point of the convex defined by (2.1) and (2.2). It can be shown that for  $n = 6$ , this is the only other kind of extreme point in addition to the tours.

3. The method. The title of this section is perhaps pretentious, as we don't have a method in any precise sense. What we do is this: Pick a tour  $x$  which looks good, and consider it as an extreme point of  $C_1$ ; use the simplex algorithm to move to an adjacent extreme point  $e$  in  $C_1$  which gives a smaller value of the functional; either  $e$  is a tour, in which case start again with this new tour, or there exists a hyperplane separating  $e$  from the convex of tours; in the latter case cut down  $C_1$  by one such hyperplane that passes through  $x$ , obtaining a new convex  $C_2$  with  $x$  as an extreme point. Starting with  $x$  again, repeat the process until a tour  $\hat{x}$  and a convex  $C_m \supset T_n$  are obtained over which  $\hat{x}$  gives a minimum of  $\sum d_{ij}x_{ij}$ .\*

The difficulties one might get into this way are manifest. It may be that a fractional extreme point is arrived at for which it is hard to find a separating hyperplane, or the process may be too tedious to be practical.\*\* In rebuttal, we can only say that the problems we have attempted have yielded rather quickly. For the 49-city problem (which was actually solved as a 42-city problem), we

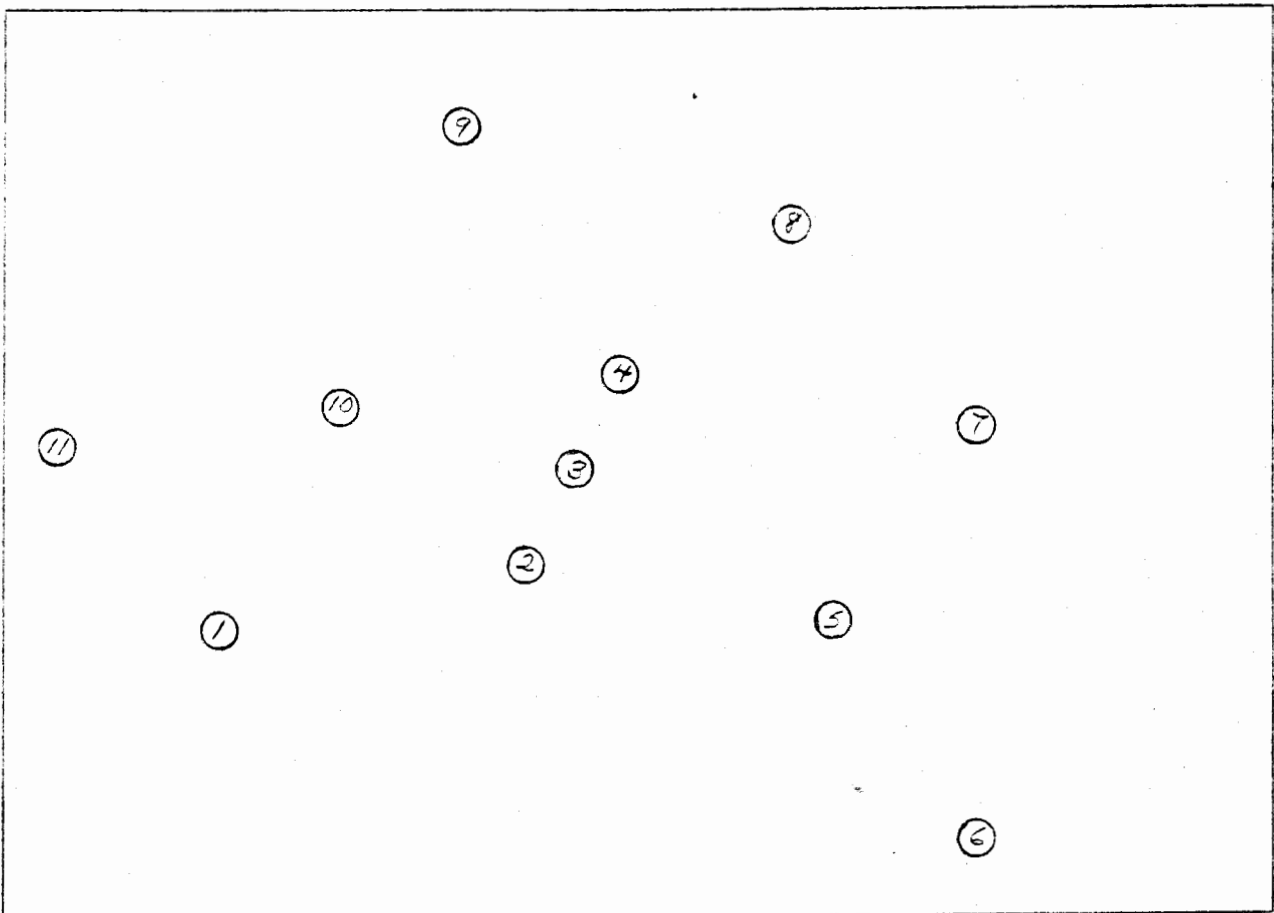
---

\*Patching up the definition of  $C$  in this way has been called the "finger in the dike" method (E. W. Paxson).

\*\*Not to mention that an infinite number of steps might be required unless one is adding extreme hyperplanes of  $T_n$ .

were lucky in guessing an optimal tour to begin with and the number of relations we required in addition to those defining  $C_1$  was 25. For other problems we have solved  $m$  was quite small also.

In the remainder of this section, we will give a description of the solution of the following 11-city problem



where

D =

1														
2	11													
3	15	5												
4	19	10	5											
5	23	13	10	10										
6	27	18	19	20	11									
7	32	22	17	14	11	19								
8	28	20	14	9	15	25	11							
9	22	18	14	11	21	32	23	13						
10	11	9	8	11	18	26	25	18	12					
11	10	17	18	21	28	34	35	28	18	11				
	1	2	3	4	5	6	7	8	9	10	11			

If one selects as a starting point the (non-optimal) tour (1, 6, 5, 7, 8, 9, 4, 3, 2, 10, 11), then the basis B for the programming problem defined by (2.1) is

B =[illegible]

The first thing to be done in the simplex algorithm is to compute the vector  $\pi$  satisfying  $\pi B = d$ , where  $d$  is the vector of distances corresponding to the tour,

$$d = (d_{61}, d_{65}, \dots, d_{11,1}) = (27, 11, 11, 11, 13, 11, 5, 5, 9, 11, 10).$$

This is easily done because of the near triangular nature of the set of equations. Let  $\pi_1$  be unknown for the moment. Then

$$\pi_6 = 27 - \pi_1$$

$$\pi_5 = -16 + \pi_1$$

$$\pi_7 = 27 - \pi_1$$

$$\pi_8 = -16 + \pi_1$$

$$\pi_9 = 29 - \pi_1$$

$$\pi_4 = -18 + \pi_1$$

$$\pi_3 = 23 - \pi_1$$

$$\pi_2 = -18 + \pi_1$$

$$\pi_{10} = 27 - \pi_1$$

$$\pi_{11} = -16 + \pi_1$$

$$2\pi_1 - 16 = 10 \quad ,$$

hence  $\pi_1 = 13$ . A quicker way is to note to begin with that

$$2\pi_1 = d_{61} - d_{65} + d_{75} - d_{87} + \dots + d_{11,1} \quad ,$$

and then compute the other components.

The next step is to compare  $\pi_i + \pi_j$  with  $d_{ij}$ . If  $\pi_i + \pi_j \leq d_{ij}$ , the tour is optimal. If not, find  $(i,j)$  for which  $\pi_i + \pi_j - d_{ij} > 0^*$

---

\*A good general rule is to choose  $\pi_i + \pi_j - d_{ij} = \max$ .

and introduce the vector  $P_{1j}$  in the basis. In the example,  $\pi_{10} + \pi_9 - d_{10,9} = 30 - 12 = 18$ , and so we bring in  $P_{10,9}$ . To do this, we need to represent  $P_{10,9}$  as a linear combination of the column vectors of  $B$ . Again this is simply done. The vector  $P_{10,9}$  has 1 in components 9 and 10, zeros elsewhere, and hence satisfies the equation

$$P_{10,9} - P_{9,8} + P_{8,7} - P_{7,5} + P_{6,5} - P_{6,1} + P_{11,1} - P_{11,10} = 0.$$

Finally, multiply this equation by  $\theta$ , add the result to the vector equation

$$P_{61} + P_{65} + P_{75} + P_{87} + P_{98} + P_{94} + P_{43} + P_{32} + P_{10,2} + P_{11,10} + P_{11,1} = \begin{pmatrix} 2 \\ \cdot \\ \cdot \\ \cdot \\ 2 \end{pmatrix},$$

and find the largest value of  $\theta$  consistent with the inequalities of (2.1),  $x_{1j} \geq 0$ . Here  $\theta = 1$ , and we have the new solution to the programming problem,

$$2P_{65} + 2P_{87} + P_{94} + P_{43} + P_{32} + P_{10,2} + P_{10,9} + 2P_{11,1} = \begin{pmatrix} 2 \\ \cdot \\ \cdot \\ \cdot \\ 2 \end{pmatrix}.$$

This is not a tour; it consists of the 5-loop (9,4,3,2,10), the isolated segments with weights 2, (65), (87), (11,1), and hence violates 4 of the relations (2.2). Select one of these relations and add it to the original problem, say

$$x_{11,1} \leq 1,$$



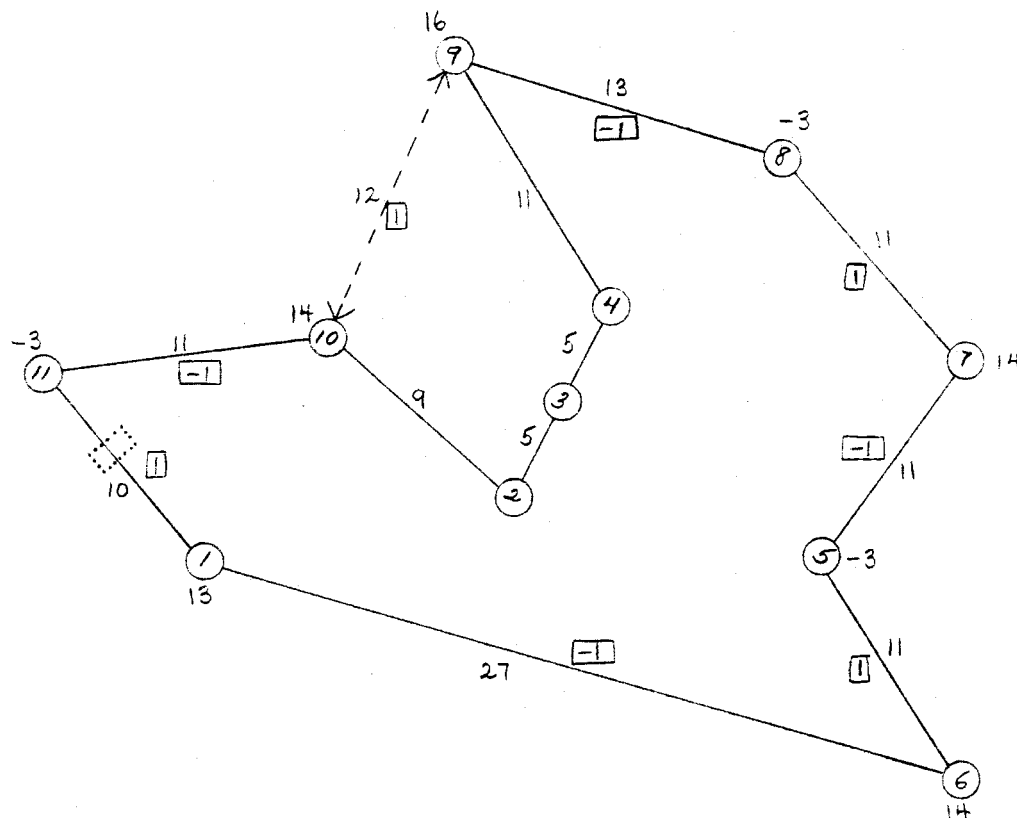
or in the programming form of equations in non-negative variables,

$$x_{11,1} + y_{11,1} = 1, \quad y_{11,1} \geq 0,$$

thus obtaining a new programming problem in which a basis  $B'$  corresponding to the tour has, in addition to the column vectors of  $B$ , the additional vector  $P_{10,9}$ . Note that these vectors now have 12 components, the 12<sup>th</sup> component being zero for all except  $P_{11,1}$  which has a 1 in the 12<sup>th</sup> position corresponding to the added equation. Also the tour is a degenerate solution to the new problem, since  $x_{10,9} = 0$ .

This discussion can be summed up in the following map:

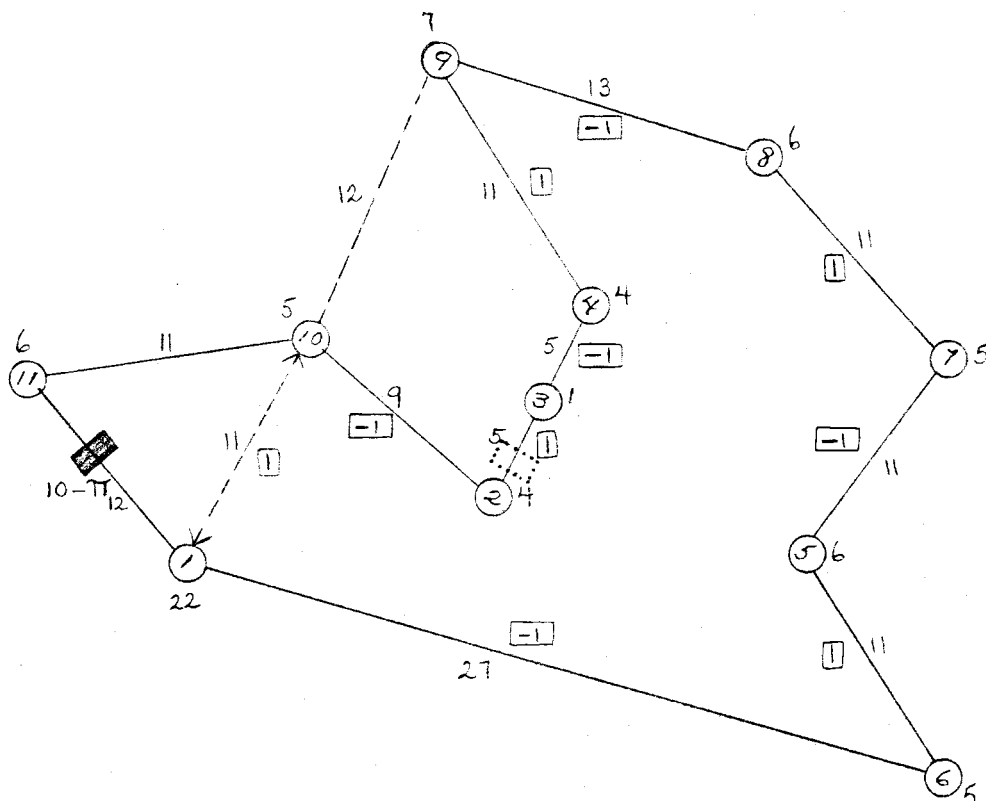
I



where the number beside each city is the value of the corresponding component of  $\pi$ , those on the segments are the distances  $d_{ij}$ . The dotted segment (10,9) represents the vector  $P_{10,9}$  to be brought into the basis and the numbers in boxes give its representation in terms of the basis. Note that to get the representation, insist that the numbers on the links emanating from any city add to zero. Consequently there are only two possibilities to consider; either the loop (10,9,4,3,2) or (10,9,8,7,5,6,1,11), the first of which must be discarded since it has an odd number of sides. The bar across (11,1) indicates that the variable  $x_{11,1}$  has an upper bound in the next programming problem.

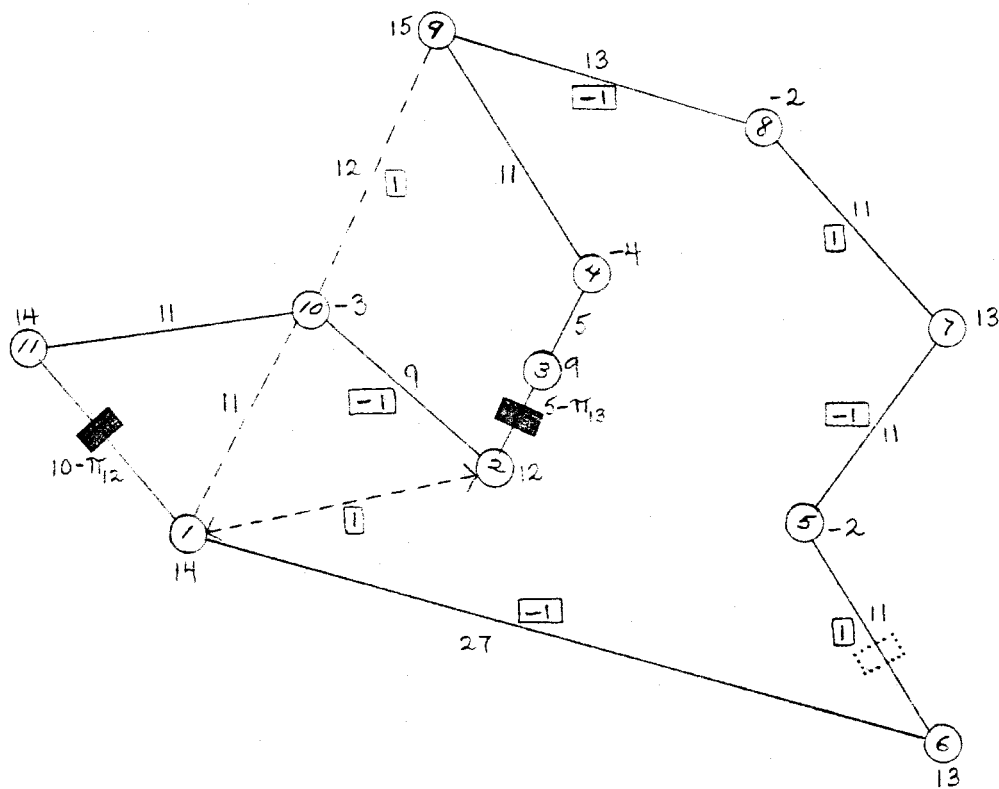
The rest of the analysis of this example will be presented as a sequence of maps.

II

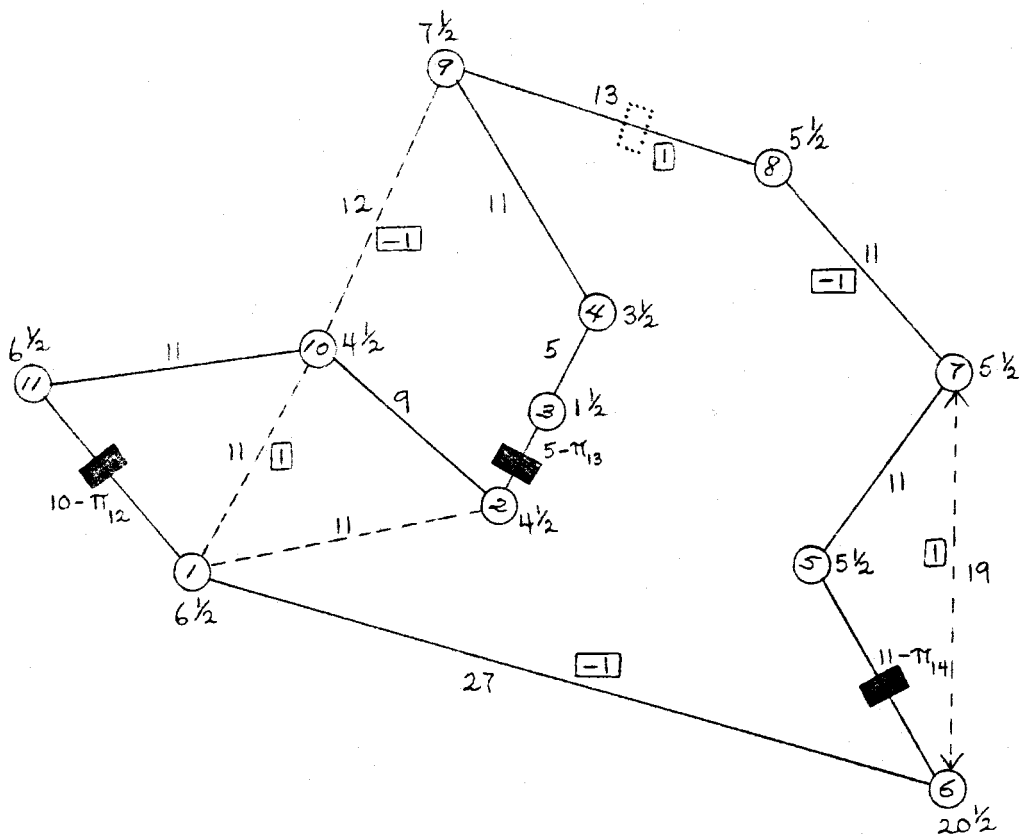


$\pi$  is computed around the loop (9,4,3,2,10), then out from 9 to 8, 7,5,6,1, from 10 to 11.  $\pi_{12}$ , the component of  $\pi$  corresponding to the added relation, is determined from  $\pi_{11} + \pi_1 - d_{11,1} = -\pi_{12}$ ,  $\pi_{12} = -18$ . To prove optimality of a tour by the simplex method, one must have these additional components of  $\pi$  negative (such as  $\pi_{12}$  in this case). Since  $\pi_{10} + \pi_1 (= 27) > d_{10,1} (= 11)$ , we introduce (10,1) into the basis. Note that in its representation, the coefficient of  $P_{11,1}$  must be zero, since the representation has to "balance" on the added relation and  $P_{11,1}$  is the only column in the basis with a non-zero component in the added relation. The resulting solution  $x$  violates  $x_{32} \leq 1$ , and we add this condition.

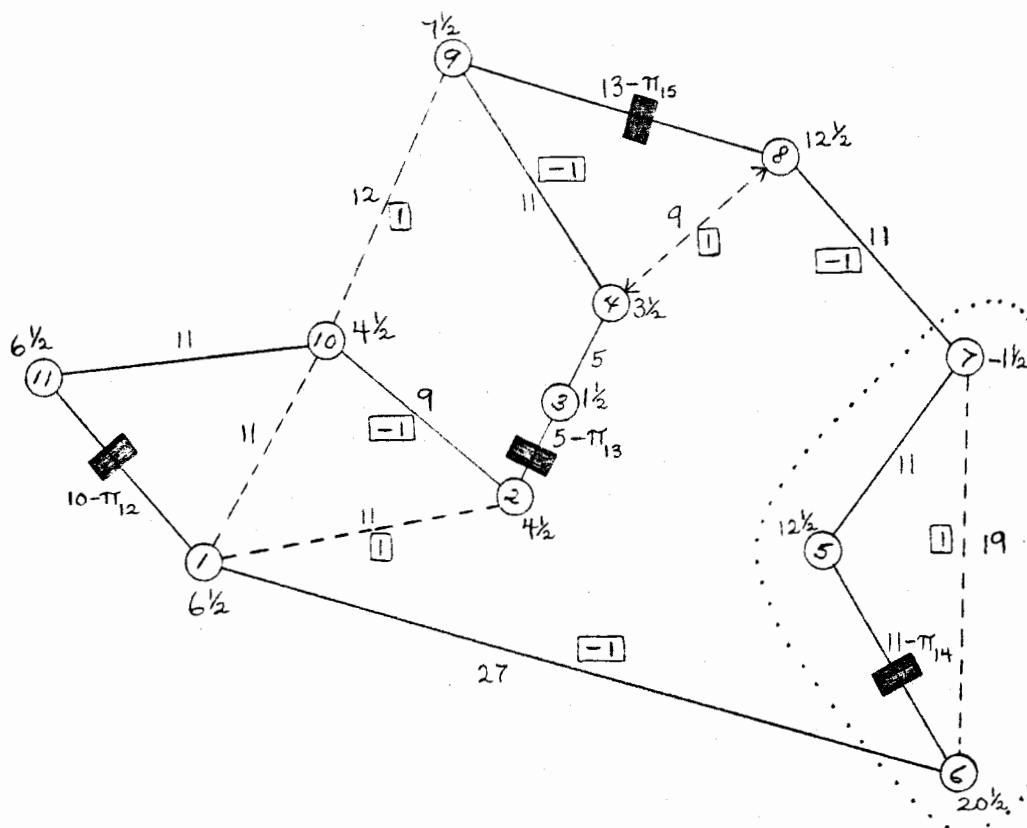
## III



## IV

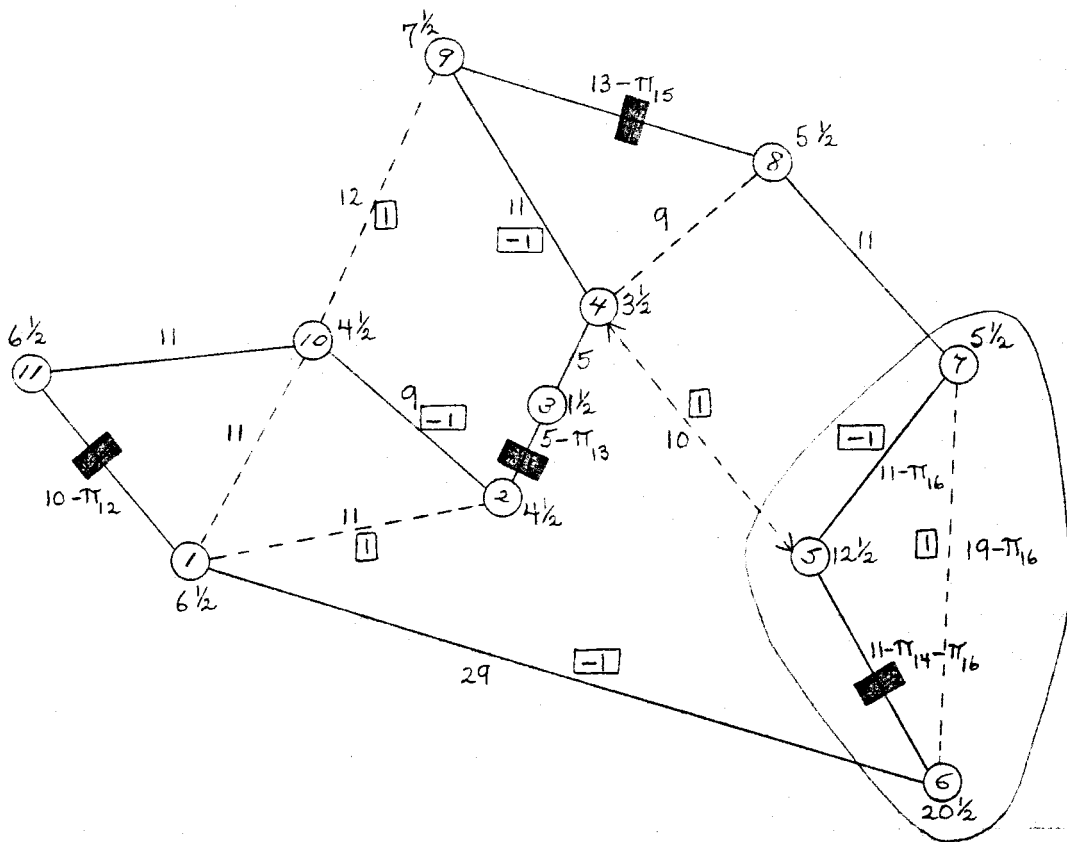


V



Here, after putting (84) in the basis, the solution consists of the loops (1,2,3,4,8,9,10,11) and (5,6,7), so we impose the condition  $x_{65} + x_{76} + x_{75} \leq 2$ , a relation of form (2.3) with  $I_1 = \{5,6,7\}$ . The imposition of this condition is indicated by the dotted loop on the map.

VI

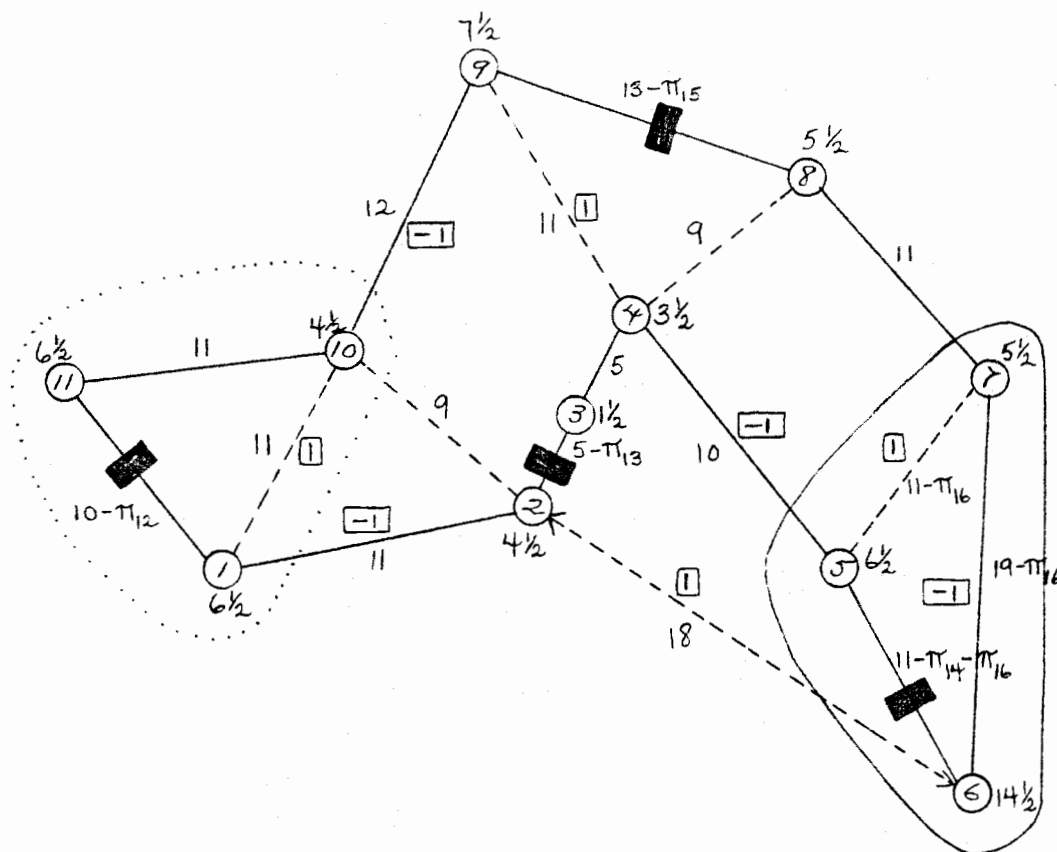


$\pi$  was computed as follows:  $\pi_1 = \frac{11-9+11}{2} = 6 \frac{1}{2}$  (from the triangle (1,2,10)), the other components by branching out from 1, i.e., 1, 10, 9, 4, 8, 7, 5; also 10, 11; 10, 2; and 1, 6. Having  $\pi_6$  and  $\pi_7$  gives  $\pi_{16}$ , hence  $\pi_{14}$ , and of course  $\pi_{12}$ ,  $\pi_{13}$ ,  $\pi_{15}$  can be solved for, e.g.,  $13 - \pi_{15} = \pi_9 + \pi_8$ . On putting (5,4) in the basis, we move to the new tour (1,2,3,..., 10,11). Note that the equation (54) - (75) + (87) - (84) = 0 would be valid except for the last components of the vectors, which correspond to the condition  $x_{75} + x_{76} + x_{65} \leq 2$ . Having arrived at this point, one could start afresh with the new tour. It seems better in practice to keep most of the conditions previously generated. Accordingly, we drop the

link (61), add the link (54), and keep all the conditions.

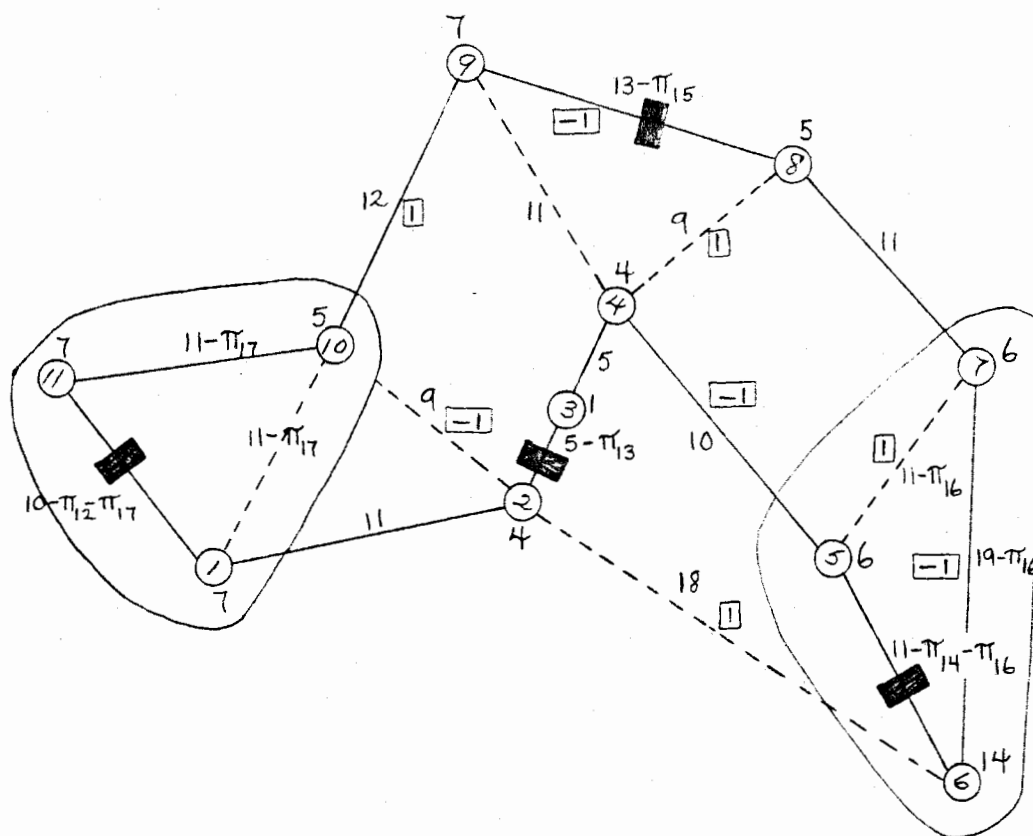
Recall from the simplex procedure also that since  $\pi_4 + \pi_5 - d_{54} = 6$  and  $\theta = 1$ , the new tour is 6 units shorter than the old tour.

VII



At this stage, the only candidate for admission to the basis is (62), for which  $\pi_6 + \pi_2 - d_{62} = 1$ . As we shall see in the next section, this implies that the tour (1,2,...,11) is within 1 unit of the minimum tour.

## VIII



For the first time one of the components of  $\pi$  corresponding to an added relation is of wrong sign,  $\pi_{15} = 1$ . Thus, the unit vector  $U_{15}$  is to be brought into the basis. The boxed numbers symbolize the equation

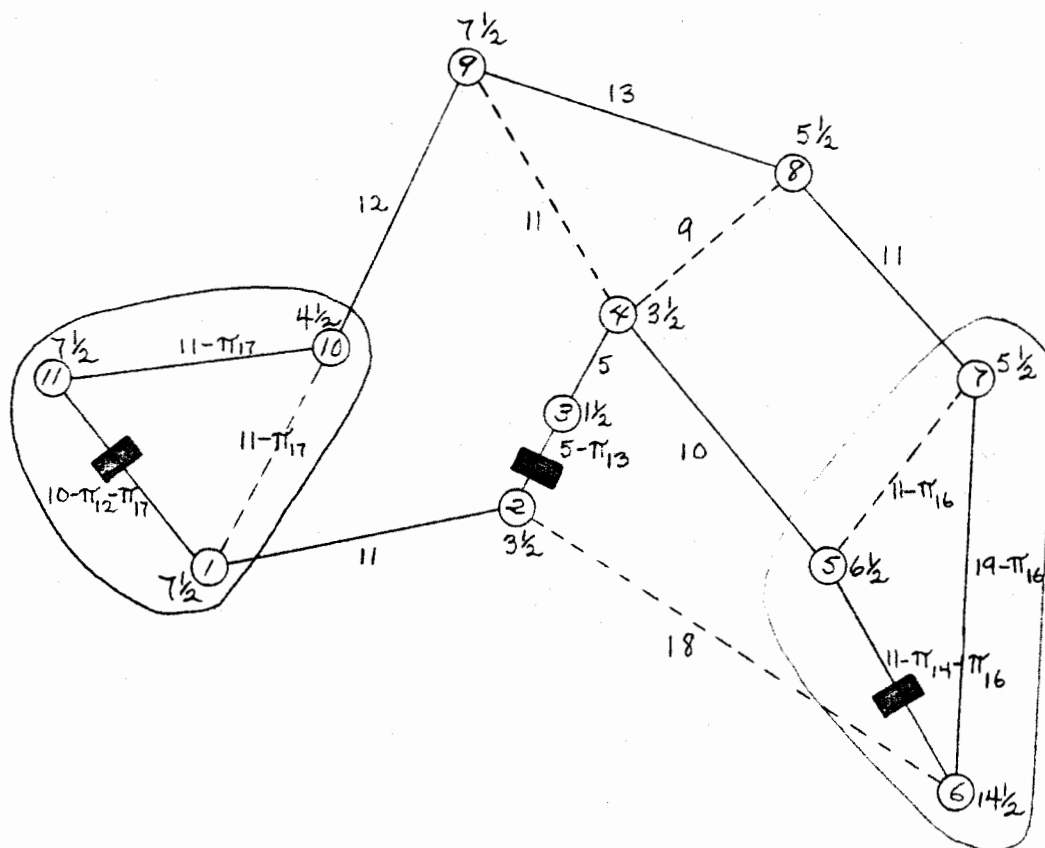
$$-(P_{98} - U_{15}) + P_{10,9} - P_{10,2} + P_{62} - P_{76} + P_{75} - P_{54} + P_{84} = 0$$

and we have  $U_{15}$  expressed as a linear combination of vectors in the basis. Observe that  $\theta = 0$  since  $x_{10,2} = 0$ . Dropping  $P_{10,2}$  from the basis and removing the restriction  $x_{98} \leq 1$  gives the



following map, which proves optimality of the tour  $(1, 2, \dots, 11)$ .

IX



4. An estimation procedure. In any linear programming problem with bounded variables, a crude estimate of how much a basic solution differs from an optimal solution can be obtained as follows: Let the programming problem be to minimize  $\sum_{i=1}^n c_i x_i$  subject to  $\sum_{i=1}^n x_i P_i = Q$ ,  $0 \leq x_i \leq r_i$ , and suppose  $\bar{x}_j$ ,  $j = 1, \dots, m$ , is a basic solution. Define  $\pi$  by  $\pi P_j = c_j$  and write  $\delta_1 = \pi P_1 - c_1$ .

For any  $x$  satisfying  $\sum_{i=1}^n x_i P_i = Q$ ,

$$(4.1) \quad \pi Q = \sum_{i=1}^n x_i \pi P_i = \sum_{j=1}^m \bar{x}_j \pi P_j = \sum_{j=1}^m \bar{x}_j c_j,$$

and hence the identity

$$(4.2) \quad \sum_{i=1}^n x_i c_i = \sum_{i=1}^n x_i \pi P_i - \sum_{i=1}^n x_i \delta_i$$

shows that

$$(4.3) \quad \sum_{i=1}^n x_i c_i = \sum_{j=1}^m \bar{x}_j c_j - \sum_{i=1}^n x_i \delta_i.$$

Thus, setting  $x_i = 0$  if  $\delta_i \leq 0$ ,  $x_i = r_i$  otherwise, one gets

$$(4.4) \quad \sum_{i=1}^n x_i c_i \geq \sum_{j=1}^m \bar{x}_j c_j - \sum_{\delta_i > 0} r_i \delta_i.$$

This estimate is usually not good enough to be of much value, but in the latter stages of solving a problem, when the  $\delta_i > 0$  are small, both in number and magnitude, it is sometimes useful. The inequality can be sharpened somewhat by maximizing  $\sum_{\delta_i > 0} x_i \delta_i$  subject to the restrictions on  $x$ . Often it is possible to do this by inspection when most of the  $\delta_i$  are negative.

For the traveling salesman problem the variables  $x_{ij}$  corresponding to weights on the segments must be either 0 or 1 if

$x$  is a tour. Hence if a number  $E$  has been determined in one of the ways described, so that  $\sum d_{ij} \bar{x}_{ij} \leq \min_x \sum d_{ij} x_{ij} + E$ , it follows from (4.3) that any tour which has weight 1 on an  $(i,j)$  link for which  $(\pi P_{ij} - d_{ij}) < -E$  can not be minimal. For example, consider map VII of the preceding section, where, as was pointed out,  $E = 1$ . Since  $\pi P_{ij} - d_{ij} < -1$  for all links except those shown on the map, an optimal tour can utilize no other links with positive weights. At this point one can proceed combinatorially. Indeed, it is easy to see that only one tour can be traced using the lines of the map, namely,  $(1,2,\dots,11)$ . Hence, it is optimal and uniquely so.

Thus, it may be that even in problems which turn out to be difficult to solve completely by sequential programming, a sufficiently close estimate can be obtained to reduce the problem to one in which the only candidates for optimal tours are readily enumerated.

5. A 49-city problem. In order to try the method on a large problem, the following set of 49 cities, one in each state and the District of Columbia, was selected:

- |                              |                            |
|------------------------------|----------------------------|
| 1. Manchester, New Hampshire | 8. Chicago, Illinois       |
| 2. Montpelier, Vermont       | 9. Milwaukee, Wisconsin    |
| 3. Detroit, Michigan         | 10. Minneapolis, Minnesota |
| 4. Cleveland, Ohio           | 11. Pierre, South Dakota   |
| 5. Charleston, West Virginia | 12. Bismarck, North Dakota |
| 6. Louisville, Kentucky      | 13. Helena, Montana        |
| 7. Indianapolis, Indiana     | 14. Seattle, Washington    |

- |                             |                              |
|-----------------------------|------------------------------|
| 15. Portland, Oregon        | 29. Dallas, Texas            |
| 16. Boise, Idaho            | 30. Little Rock, Arkansas    |
| 17. Salt Lake City, Utah    | 31. Memphis, Tennessee       |
| 18. Carson City, Nevada     | 32. Jackson, Mississippi     |
| 19. Los Angeles, California | 33. New Orleans, Louisiana   |
| 20. Phoenix, Arizona        | 34. Birmingham, Alabama      |
| 21. Santa Fe, New Mexico    | 35. Atlanta, Georgia         |
| 22. Denver, Colorado        | 36. Jacksonville, Florida    |
| 23. Cheyenne, Wyoming       | 37. Columbia, South Carolina |
| 24. Omaha, Nebraska         | 38. Raleigh, North Carolina  |
| 25. Des Moines, Iowa        | 39. Richmond, Virginia       |
| 26. Kansas City, Missouri   | 40. Washington, D. C.        |
| 27. Topeka, Kansas          | 41. Boston, Massachusetts    |
| 28. Oklahoma City, Oklahoma | 42. Portland, Maine          |

- A. Baltimore, Maryland
- B. Wilmington, Delaware
- C. Philadelphia, Pennsylvania
- D. Newark, New Jersey
- E. New York, New York
- F. Hartford, Connecticut
- G. Providence, Rhode Island

The reason for picking this particular set was that most of the road distances between them were easy to get from a Rand McNally atlas. The array on page 27, which is part of the original one prepared by Bernice Brown, gives  $d_{ij} = \frac{d'_{ij} - 11}{17}$ , \*  $i, j = 1, 2, \dots, 42$ ,

---

\*This particular transformation was chosen to make the  $d'_{ij}$  of the original table less than 256 which would permit a compact storage of the distance table in binary representation. (No use was made of this, however.)

where  $d'_{ij}$  is the road distance in miles between  $i$  and  $j$ . The  $d_{ij}$  have been rounded to the nearest integer. Certainly such a linear transformation does not alter the ordering of the tours, although, of course, the rounding may to some extent.

We will show that the tour (see map X)

$$\tau_0 = (1, 2, \dots, 42)$$

is minimal for the subset of numbered cities. The justification for omitting the lettered cities from the formal analysis is that in driving from 40 to 41, one goes through A, B, ..., G successively.\*

The proof that  $\tau_0$  is minimal is contained essentially in map XI. To make the correspondence between XI and its programming problem clear, we will write down explicitly a set of 67 relations in non-negative variables which define a convex  $C \supset T_{42}$  and over which  $\tau_0$  affords a minimum of the functional  $\sum d_{ij}x_{ij}$ .

We distinguish the following subsets of the 42 cities:

---

\*It is important to note that this tour is optimal for the table of road distances. J. D. Williams has pointed out that there are shorter airline distance tours through these same cities.

### Table of Road Distances between Cities in Adjusted Units \*

$$\begin{aligned}
I_1 &= \{1, 2, 41, 42\} \\
I_2 &= C(I_1) \text{ (complement)} \\
I_3 &= \{3, 4, \dots, 9\} \\
I_4 &= C(I_3) \\
I_5 &= \{1, 2, \dots, 9, 29, 30, \dots, 42\} \\
I_6 &= C(I_5) \\
I_7 &= \{11, 12, \dots, 23\} \\
I_8 &= C(I_7) \\
I_9 &= \{13, 14, \dots, 23\} \\
I_{10} &= C(I_9) \\
I_{11} &= \{13, 14, 15, 16, 17\} \\
I_{12} &= \{24, 25, 26, 27\} .
\end{aligned}$$

Except for two inequalities which we will discuss in a moment, the programming problem may now be written as (recall the notation of §2):  $\sum_{i \text{ or } j = i_0} x_{ij} = 2$  ( $i_0 = 1, \dots, 42$ ),  $x_{41,1} \leq 1$ ,  $x_{4,3} \leq 1$ ,  $x_{7,6} \leq 1$ ,  $x_{9,8} \leq 1$ ,  $x_{12,11} \leq 1$ ,  $x_{14,13} \leq 1$ ,  $x_{15,14} \leq 1$ ,  $x_{20,19} \leq 1$ ,  $x_{23,22} \leq 1$ ,  $x_{25,24} \leq 1$ ,  $x_{27,26} \leq 1$ ,  $x_{29,28} \leq 1$ ,  $x_{31,30} \leq 1$ ,  $x_{33,32} \leq 1$ ,  $x_{35,34} \leq 1$ ,  $x_{37,36} \leq 1$ ,  $\sum_{1,2} \geq 2$ ,  $\sum_{3,4} \geq 2$ ,  $\sum_{5,6} \geq 2$ ,  $\sum_{7,8} \geq 2$ ,  $\sum_{9,10} \geq 2$ ,  $\sum_{11} \leq 4$ ,  $\sum_{12} \leq 3$ . The remaining two relations (66 and 67) are perhaps most easily described verbally. The first says that  $x_{14,15}$  minus the sum of all other  $x_{ij}$  on segments out of 15, 16, 19, except for  $x_{18,15}$ ,  $x_{18,16}$ ,  $x_{17,16}$ ,  $x_{19,18}$ , and  $x_{20,19}$ , which do not appear in the sum, is not positive; the second that  $\sum a_{ij} x_{ij} \leq 42$ , where  $a_{23,22} = 2$ ,  $a_{26,25} = 0$ ,  $a_{ij} = 0$  if  $(i,j)$  is a link between  $I_5$  and  $I_6$  except that  $a_{109} = a_{29,28} = 1$ , also  $a_{ij} = 0$

if either  $i$  or  $j$  is one of 10, 21, 25, 26, 27, 28 except that if  $(i,j) \neq (26,25)$  appears as a link of map XI (i.e.,  $P_{ij}$  is in the basis), then  $a_{ij} = 1$ , and  $a_{ij} = 1$  for all other  $a_{ij}$  not determined above.\* These two inequalities are satisfied by all tours. For example, if a tour were to violate the first one, it must have successively  $x_{15,14} = 1$ ,  $x_{18,15} = 1$ ,  $x_{18,16} = 1$ , but also  $x_{19,18} = 1$ , a contradiction. The argument that each tour satisfies the second inequality is similar but somewhat more involved, and we omit it. These relations were imposed to cut out fractional extreme points which satisfied all the conditions (2.1) and (2.2).

We assert that  $\pi$ , as given in XI, satisfies  $\pi \bar{P}_{ij} = d_{ij}$  for  $\bar{P}_{ij}$  in the basis,  $\pi P_{ij} \leq d_{ij}$  for  $P_{ij}$  not in the basis, and  $\pi_{43}, \dots, \pi_{67}$  are appropriately positive or negative (positive if the corresponding added inequality has been written in the form  $\sum a_{ij}x_{ij} \geq r$ , negative if  $\sum a_{ij}x_{ij} \leq r$ ), with this exception:  $\pi_{52} = 1/2$ . (The variable in this case has an upper bound of 1 and was introduced to take care of the restraint  $x_{24,25} \leq 1$ .) This proves, since  $E = 1/2$  and all the  $d_{ij}$  are integers, that  $\tau_0$  is minimal. The length of  $\tau_0$  is 699 units, or 12,345 miles except for rounding errors.

It can be shown by introducing all links for which  $\pi P_{ij} - d_{ij} \geq -1/2$  that  $\tau_0$  is the unique minimum. There are only 7 such links in addition to those shown in XI, and consequently all

---

\*We are indebted to I. Glicksberg for pointing out relations of this kind to us.

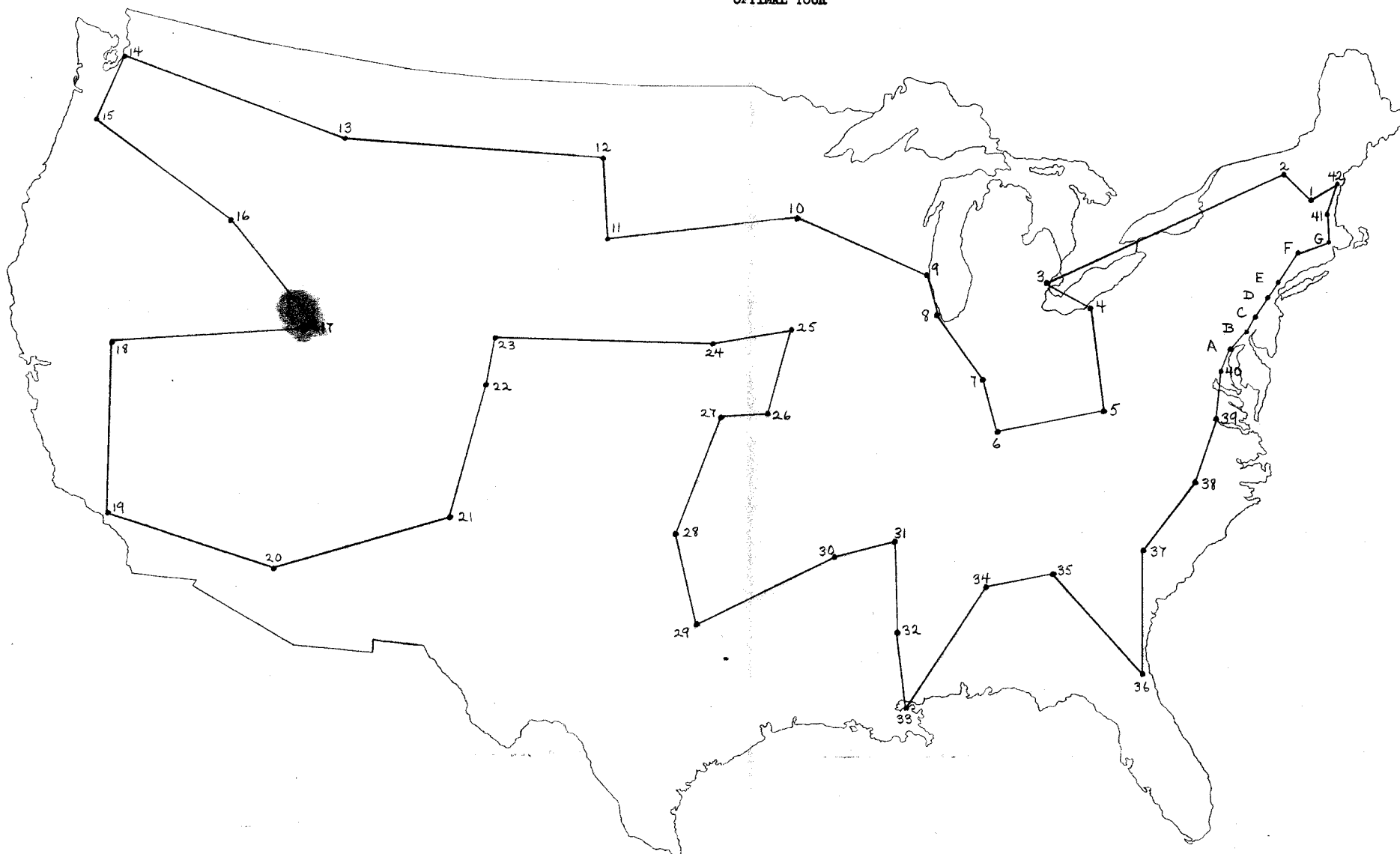


possible tying tours were enumerated without too much trouble. None of them proved to be as good as  $\tau_0$ .

6. Concluding remark. It is clear that we have left unanswered practically any question one might pose of a theoretical nature concerning the traveling salesman problem; however, we hope that the feasibility of attacking problems involving a moderate number of points has been successfully demonstrated, and that perhaps some of the ideas can be used in problems of similar nature.

X

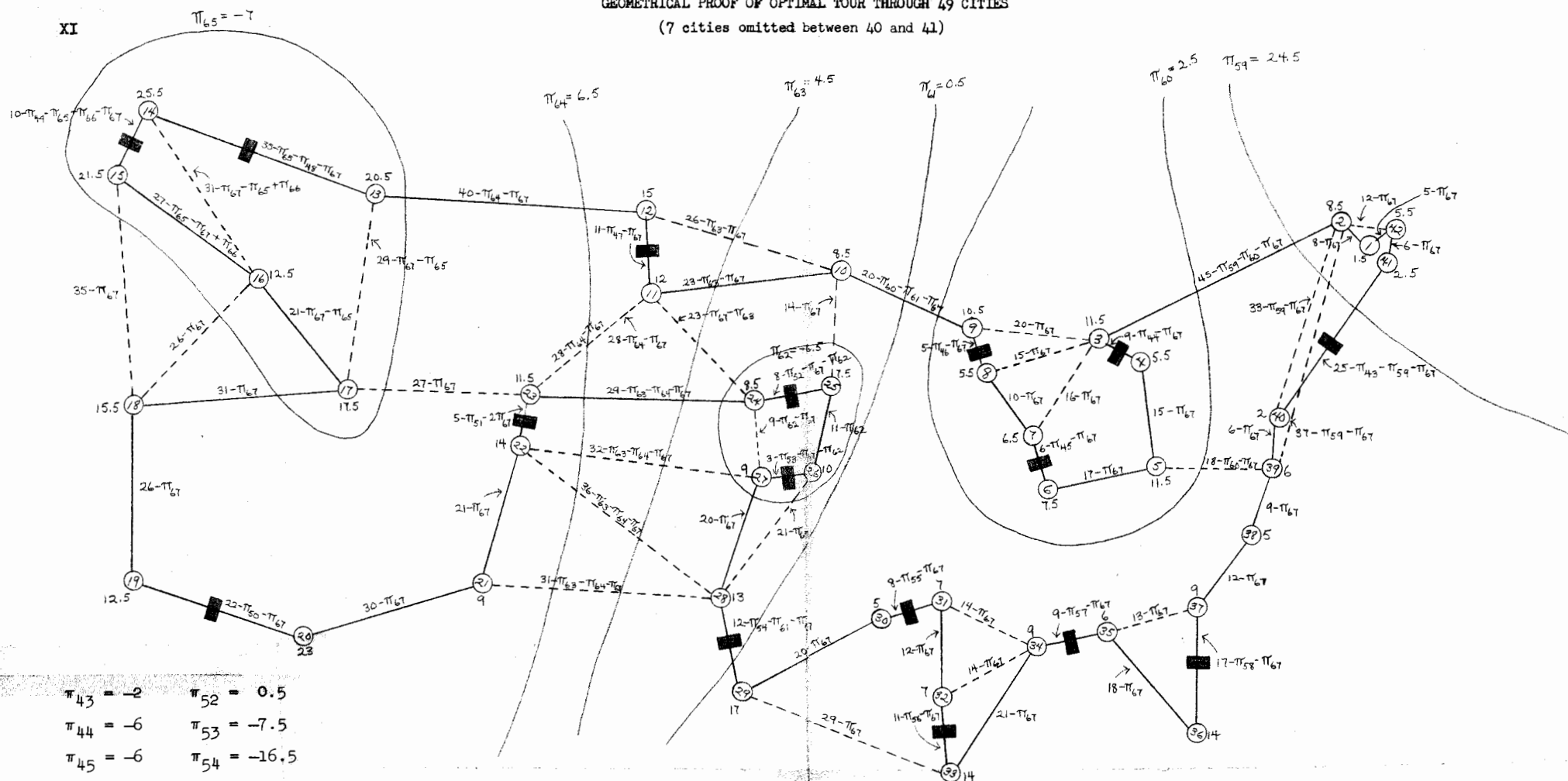
OPTIMAL TOUR \*



\*This tour has a length of 12,345 miles when the adjusted units are expressed in miles.

GEOMETRICAL PROOF OF OPTIMAL TOUR THROUGH 49 CITIES  
(7 cities omitted between 40 and 41)

XI



$\pi_{43} = -2$	$\pi_{52} = 0.5$
$\pi_{44} = -6$	$\pi_{53} = -7.5$
$\pi_{45} = -6$	$\pi_{54} = -16.5$
$\pi_{46} = -9$	$\pi_{55} = -2$
$\pi_{47} = -14$	$\pi_{56} = -8$
$\pi_{48} = -2$	$\pi_{57} = -4$
$\pi_{49} = -26$	$\pi_{58} = -4$
$\pi_{50} = -11.5$	$\pi_{66} = -2$
$\pi_{51} = -16.5$	$\pi_{67} = -2$

**Note:** In comparing  $\pi P_{ij}$  with  $d_{ij}$ , remember that the components of  $P_{ij}$ , corresponding to the added relations must be considered. The added relations are pictured graphically above except for the last two, 66 and 67. A loop drawn about a subset of points corresponds to a relation of the kind  $\sum_1 \leq n_1 - 1$ ; an arc separating a set of points from its complement corresponds to  $\sum_{1,j} \geq 2$ . For example, to compare  $\pi P_{25,9}$  with  $d_{25,9} = 21$ , observe that the segment (25,9) crosses two arcs, i.e.,  $P_{25,9}$  has a 1 in components 60 and 61; it has a zero in components 66 and 67 since  $x_{25,9}$  does not appear in those relations, and of course, has zeros elsewhere, except for components 25 and 9. Hence  $\pi P_{25,9} = \pi_{25} + \pi_9 + \pi_{60} + \pi_{61} = 7.5 + 10.5 + 2.5 + 0.5 = 21 \leq d_{25,9}$ . One more example:  $\pi P_{27,25} = \pi_{27} + \pi_{25} + \pi_{62}$ , since (27,25) is contained within the loop representing the condition  $\sum_{12} \leq 3$ .

## REFERENCES

- [1] Dantzig, G. B., Orden, A., and Wolfe, P., "The Generalized Simplex Method for Minimizing a Linear Form under Linear Inequality Restraints," RAND P-392, 5 April 1954.
- [2] ———, "Application of the Simplex Method to a Transportation Problem," Activity Analysis of Production and Allocation, RAND Report R-193, June, 1951.
- [3] Heller, I., "On the Problem of Shortest Path Between Points," I and II (abstract), Bulletin of the American Mathematical Society, Vol. 59, No. 6, November, 1953.
- [4] Kuhn, H. W., "The Traveling Salesman Problem," to appear in the Proceedings of the Sixth Symposium in Applied Mathematics of the American Mathematical Society, McGraw-Hill Book Company.
- [5] Votaw, D. F., and Orden, A., "Personnel Assignment Problem," Symposium on Linear Inequalities and Programming, June 14-16, 1951, Comptroller, Headquarters U. S. Air Force.
- [6] von Neumann, J., "A Certain Zero-sum Two-person Game Equivalent to the Optimal Assignment Problem," Contributions to the Theory of Games II, Princeton University Press, 1953.
- [7] Ball, W. W. R., Mathematical Recreations and Essays, as Rev. by H. S. M. Coxeter, Eleventh edition, 1939, Macmillan, New York.
- [8] Tutte, W. T., "On Hamiltonian Circuits," London Mathematical Society Journal, Vol. XXI, Part 2, April 1946, No. 82, page 98-101.
- [9] Verblunsky, S., "On the Shortest Path Through a Number of Points," Proceedings of the American Mathematical Society, Vol. II, No. 6, December 1951.