
A Method for Solving Traveling-Salesman Problems

Author(s): G. A. Croes

Source: *Operations Research*, Nov. - Dec., 1958, Vol. 6, No. 6 (Nov. - Dec., 1958), pp. 791-812

Published by: INFORMS

Stable URL: <https://www.jstor.org/stable/167074>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/167074?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Operations Research*

JSTOR

Operations Research

November - December 1958

A METHOD FOR SOLVING TRAVELING-SALESMAN PROBLEMS*

G. A. Croes†

*Exploration and Production Research Division, Shell Development
Company, Houston, Texas*

(Received June 20, 1957)

The traveling-salesman problem is a generalized form of the simple problem to find the smallest closed loop that connects a number of points in a plane. Efforts in the past to find an efficient method for solving it have met with only partial success. The present paper describes a method of solution that has the following properties: (a) It is applicable to both symmetric and asymmetric problems with random elements. (b) It does not use subjective decisions, so that it can be completely mechanized. (c) It is appreciably faster than any other method proposed. (d) It can be terminated at any point where the *solution* obtained so far is deemed sufficiently accurate.

THE TRAVELING-SALESMAN problem got its name from the following simple situation: A traveling salesman leaves his home town in the morning, makes the round of all the towns in a set of towns, and

* NOTE BY THE EDITOR: This paper was received from G. A. Croes shortly before he returned to Holland after an assignment at Shell Development Company. Subsequent attempts by the Editor to contact Croes have failed. As a result, the author has neither seen proof of this article nor has he taken advantage of the constructive comments of the referees. Minor corrections suggested by the referees have been made. One comment that should be called to the reader's attention is that the use of x_{ij} with i representing the *column* and j the *row* is the opposite of the usual usage.

† The author is indebted to H. RAINBOW for his active interest in the investigation, resulting in a number of valuable suggestions, and to A. GINZBURG, who suggested some important improvements in the presentation of the subject. The author also wishes to express his appreciation to Shell Development Company for permission to publish this article.

arrives back home at nightfall. In which sequence should he visit the towns in order to make the total distance traveled as small as possible?

A convenient way to display the data of the above problem (the distance between each pair of towns in the set) is in the form of the triangular distance table, well known from atlases and roadmaps. These arrays are in principle square, symmetrical matrices with the diagonal and symmetrical elements left out. In more sophisticated forms of the problem, these matrices can be asymmetric.

The problem is one of the few of the class of programming problems for which no completely satisfactory method of solution exists. The only fairly successful attempt at finding such a method (for symmetrical problems) has been reported by DANTZIG ET AL.^[1] The method they propose is to formulate the problem as a problem in linear programming, consider the majority of the necessary restraints as not restricting the solution, and solve the problem under the set of remaining restraints by means of the standard simplex method. A partial justification for their selection of the restraints to be included in the calculation has been given by HELLER.^[2] A review of our present knowledge of the problem has been given by FLOOD.^[3]

The most general statement of the problem is as follows: Given the matrix $C = \| c_{ij} \|$ ($i=1, 2, \dots, n; j=1, 2, \dots, n$) and the set S of all possible sequences s of ordered pairs (i, j) , for which:

1. All values of $i=1, 2, \dots, n$ and $j=1, 2, \dots, n$ occur at least once in every s .
2. If (i_1, j_1) and (i_2, j_2) are consecutive pairs of s , then $j_1 = i_2$.
3. If (i_1, j_1) is the first element and (i_m, j_m) is the last element of s , then $j_m = i_1$.

Find the sequence $s_0 \in S$ for which $M(s_0) = \min_s M(s)$, where the measure $M(s)$ of the sequence s is defined by $M(s) = \sum_{(i,j) \in s} c_{ij}$.

The set S can be represented in two others ways: (a) By the set T of sequences t of integers, in which every integer between 1 and n occurs at least once in every t . (b) By the set X of all matrices $x = \| x_{ij} \|$, which are such that if $(i, j) \in s$ then $x_{ij} = 1$ and if $(i, j) \notin s$ then $x_{ij} = 0$.

The emphasis of this report is on the solution of the general problem under two restrictions, viz:

1. In every s all values of i and j occur once and only once.
2. In the matrix C , $c_{ij} = c_{ji}$, i. e., C is a symmetrical matrix.

We deal thus especially with symmetrical problems in which 'every town is visited only once.' For this type of problem $\sum_i x_{ij} = \sum_j x_{ij} = 1$. In the final section, the method given for symmetrical problems is shown to be applicable also to asymmetrical problems, though it then requires more work.

The case where 'the towns can be visited more than once,' though of some practical importance, will not be considered in this paper.

SYMMETRICAL PROBLEMS

THE PRESENT method for solving traveling-salesman problems proceeds in three stages:

1. Find a trial solution $s \in S$, for which $M(s)$ is as small as we can make it at a first try.
2. Apply simple transformations, called 'inversions,' which transform this trial solution into other elements of S , whose measures are progressively smaller.
3. Check C for elements which might be included in the final s at an advantage. If there are any such elements try to find a transformation which decreases the measure of the sequence.

The method is elucidated by means of a problem of order 20. Smaller problems are not likely to show all the complications that might arise in the solution of large-scale problems. The method depends on the truth of a number of theorems. Proofs that were considered trivial have been omitted.

The Trial Solution

The first step in finding a trial solution is to construct a matrix $\|x_{ij}\|$ such that x_{ij} is either 1 or 0, $x_{ii}=0$, $\sum_i x_{ij} = \sum_j x_{ij} = 1$. To that end we select the smallest element of C , say $c_{i_1 j_1}$, and put $x_{i_1 j_1} = 1$. In a symmetrical problem this procedure is ambiguous, since we can choose between two symmetrical elements. We shall adopt the convention that if such a choice is indicated, we always choose the element for which $i < j$.

As $\sum_i x_{ij} = \sum_j x_{ij} = 1$, it follows that if $x_{i_1 j_1} = 1$ and $i \neq i_1$, $j \neq j_1$, then $x_{i j_1} = x_{i_1 j} = 0$. As further x_{ij} and x_{ji} cannot be elements of the same s , $x_{j_1 i_1} = 0$.

The next nonzero element of $\|x_{ij}\|$ is now chosen by selecting the smallest of the elements of C which are not ruled ineligible by the inclusion of $x_{i_1 j_1}$. This procedure is repeated until every row and column of $\|x_{ij}\|$ contains one nonzero element.

This matrix $\|x_{ij}\|$ is not yet necessarily an element of X , as it might represent a number of cyclic subsequences instead of a single cyclic sequence. In order to make it an element of X , we proceed as follows: Starting with any of the elements $x_{ij} = 1$, say $x_{i_1 j_1}$, construct the sequence $x_{i_1 j_1}, x_{i_2 j_2}, \dots, x_{i_m j_m}$, where all x_{ij} 's are 1, $j_a = i_{a+1}$ and $j_m = i_1$. If this sequence contains all nonzero elements of $\|x_{ij}\|$, this matrix is an element of X . If that is not the case, start with any unit element not contained in the previous sequence and construct a second sequence. If necessary

construct a third, fourth, etc., until all nonzero elements of $\|x_{ij}\|$ are contained in a sequence. Two of these sequences can be combined into a single sequence by taking one element out of each, say x_{ij} and x_{kl} , and replacing them by x_{il} and x_{jk} . This replacement is feasible only if x_{ii} and x_{jj} are zero and $i=l, j=k$.

It is sometimes advantageous to choose c_{ij} and c_{kl} as large as possible. Though this particular choice is arbitrary, it will be adopted as a working rule.

When the above procedure is repeated, any number of sequences can be linked together into a single sequence. The $\|x_{ij}\|$ corresponding to this single sequence is an element of X and forms our first trial solution.

For an example of the procedure see Fig. 1, where C is given as a symmetrical 20×20 matrix with random elements. The selection of the unit elements of $\|x_{ij}\|$ was made in the following sequence (the corresponding elements of C have been underlined with a dotted line): (5,9), (7,13), (13,4)(18,20), (8,16)(9,12), (1,6)(11,14), (14,10)(17,11), (4,1), (10,18), (6,17), (15,7), (16,15), (19,5)(20,3), (12,19), (2,8), (3,2), giving the two cyclic subsequence: 1,6,17,11,14,10,18,20,3,2,8,16,15,7,13,4 and 5,9,12,19. These can be combined by replacing (3,2) and (12,19) by (12,2) and (3,19), resulting in the single cyclic sequence (underlined with a solid line): 1,6,17,11,14,10,18,20,3,19,5,9,12,2,8,16,15,7,13,4.

For the purpose of demonstration the element (12,1) was ignored in determining $\|x_{ij}\|$. Inclusion of this element would have resulted immediately in a single sequence.

Inversions

Now that we have a trial solution, we shall attempt to derive from it another one, which has a smaller measure. The total number of such possible 'transformations' is of the order of $\frac{1}{2}(n-1)!$; hence a direct check on all transformations to determine whether they reduce the measure of the trial solution or not is out of the question for even a moderate value of n .

Particular transformations exist however which are both effective and easy to handle computationally. One of these transformations, say $I(i_1, i_k)$, transforms the trial solution: $1, 2, \dots, i_0, i_1, i_2, \dots, i_k, i_{k+1}, \dots, n$ into another one: $1, 2, \dots, i_0, i_k, i_{k-1}, \dots, i_2, i_1, i_{k+1}, \dots, n$. These transformations have been called 'inversions,' as they invert the sequence of integers in a part of the trial solution.

The decrease in measure caused by operating with $I(i_1, i_k)$ on the trial solution is for a symmetrical C given by

$$\Delta M(i_1, i_k) = c_{i_0, i_1} + c_{i_k, i_{k+1}} - c_{i_0, i_k} - c_{i_1, i_{k+1}}.$$

If $\Delta M > 0$, we call I 'desirable.'

1	x	29	41	<u>09</u>	18	06	42	48	74	43	51	07	36	93	58	11	51	61	30	44
2	29	x	<u>72</u>	72	50	39	60	34	25	46	25	<u>35</u>	14	20	35	83	27	86	95	30
3	41	72	x	70	54	35	59	88	19	72	87	38	24	68	63	80	58	40	89	<u>24</u>
4	09	72	70	x	60	20	24	73	79	51	43	<u>58</u>	<u>04</u>	47	29	22	48	27	88	91
5	18	50	54	60	x	17	74	93	00	76	30	55	<u>84</u>	42	47	91	21	59	<u>24</u>	80
6	<u>06</u>	39	35	20	17	x	26	60	32	63	84	21	26	96	75	14	13	51	16	83
7	42	60	59	24	74	26	x	97	65	64	13	23	03	78	<u>15</u>	30	56	22	13	58
8	48	<u>34</u>	88	73	93	60	97	x	63	27	42	62	32	20	26	05	80	52	47	36
9	74	25	19	79	<u>00</u>	32	65	63	x	71	91	05	85	51	72	53	08	49	90	39
10	43	46	72	51	76	63	64	27	71	x	66	30	57	<u>08</u>	71	19	25	10	83	40
11	51	25	87	43	30	84	13	42	91	66	x	09	26	06	99	33	<u>08</u>	99	92	31
12	07	35	38	58	55	21	23	62	<u>05</u>	30	09	x	86	27	34	72	45	59	32	77
13	36	14	24	04	84	26	<u>03</u>	32	85	57	26	86	x	12	28	24	60	19	12	20
14	93	20	68	47	42	96	78	20	51	08	<u>06</u>	27	12	x	19	77	14	22	54	77
15	58	35	63	29	47	75	15	26	72	71	99	34	28	19	x	<u>22</u>	75	28	72	64
16	11	83	80	22	91	14	30	<u>5</u>	53	19	33	72	24	77	22	x	62	79	97	47
17	51	27	58	48	21	<u>13</u>	56	80	08	25	08	45	60	14	75	62	x	91	59	75
18	61	86	40	27	59	51	22	52	49	<u>10</u>	99	59	19	22	28	79	91	x	87	04
19	30	95	<u>89</u>	88	24	16	13	47	90	83	92	<u>32</u>	12	54	72	97	59	87	x	32
20	44	30	24	91	80	83	58	36	39	40	31	77	20	77	64	47	75	<u>04</u>	32	x
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Subsequences:

1, 6, 17, 11, 14, 10, 18, 20, 3, 2, 8, 16, 15, 7, 13, 4 and
5, 9, 12, 19

Replace (3,2) and (12,19) by (3,19) and (12,2)

First Trial Solution: $M = 324$

1, 6, 17, 11, 14, 10, 18, 20, 3, 19, 5, 9, 12, 2, 8, 16, 15, 7,
13, 4.

.... = First $\parallel x_{ij} \parallel$, — = First Trial Solution

Figure 1

In the matrix as originally given, the calculation of ΔM is rather unwieldy, but it can be simplified considerably by rearranging the rows and the columns of C such that the trial solution becomes the 'slant' of the new matrix. If the rows and columns in a matrix are numbered consecutively

1, 2, \dots , n , the slant is defined to be the set $[(i, i+1)]$ ($i=1, 2, \dots, n-1$) together with the element $(n, 1)$. Another set of elements that will be used later is the 'contra-slant,' defined as the set $[(i, i-1)]$ ($i=2, 3, \dots, n$) together with the element $(1, n)$. In the new matrix the sequence of the rows and of the columns, when numbered as in the original C , is equal to that of the trial solution, which fact facilitates the rearranging procedure.

A check on which of the possible inversions is desirable, if any, can be carried out easily in the new matrix. For a symmetrical C only those inversions for which $i_1 < i_k$ have to be checked, since for such a matrix $\Delta M(i, j) = \Delta M(j+1, i-1)$.

Turning again to our example, we see that in Fig. 2 the rows and the columns of the original matrix have been rearranged to get the trial solution in the slant. A check on the possible inversions shows that the following are desirable: $I(2, 9)$, $I(3, 9)$, $I(10, 12)$, $I(10, 14)$, $I(10, 17)$, $I(10, 18)$, $I(5, 13)$, and $I(14, 18)$.

It often happens that the sequence in which simultaneously desirable inversions are to be carried out is significant. This is the case if for two desirable inversions $I(i_1, j_1)$ and $I(i_2, j_2)$: $i_1 = i_2$, $i_1 = j_2$, $j_1 = j_2$ or $i_1 < i_2 < j_1 < j_2$ (i.e., if the inversions 'have a point in common'). If one of the equalities holds, we simply select that one of the two that gives the greater decrease in measure. If the inequality holds, we have to check which one of the two possible sequences or of the two inversions alone gives the greatest gain.

Keeping in mind that for a symmetrical matrix $\Delta M(i, i) = \Delta M(j+1, i-1)$, we see that the set of desirable inversions in our example falls apart in two subsets, for each of which $i_1 = i_2$: the subset consisting of the first six inversions and that consisting of the last two inversions.

The ΔM 's of these inversions have been tabulated in Fig. 2 to the right of the matrix. It is seen that the two noninteracting inversions $I(10, 12)$ and $I(5, 13)$ give the largest possible decrease in measure. Carrying them out results in the new trial solution: 1, 6, 17, 11, 12, 19, 5, 9, 3, 20, 18, 10, 14, 2, 8, 16, 15, 7, 13, 4.

The rearranged matrix for this solution is given in Fig. 3. A check on the inversions shows that there is one desirable inversion: $I(2, 5)$. This inversion ultimately results in Fig. 4, a rearranged matrix C in which no further inversions are desirable.

Final Adjustment

The trial solution obtained in the previous section is generally close to the actual solution, and hence it might suffice for all but the more demanding practical problems. Nevertheless, a method will be given to derive the optimal sequence from the trial solution obtained by means of inversions.

We shall try to find a transformation that replaces a number of elements $c_{i,i+1}$ of the trial solution by other elements $c_{i,j}$ ($j \neq i+1$), such that (a) the sum of the new elements $\sum_N c_{ij}$ is smaller than the sum of the old elements $\sum_O c_{i,i+1}$, or $\sum_N c_{ij} - \sum_O c_{i,i+1} < 0$, (b) the remaining ele-

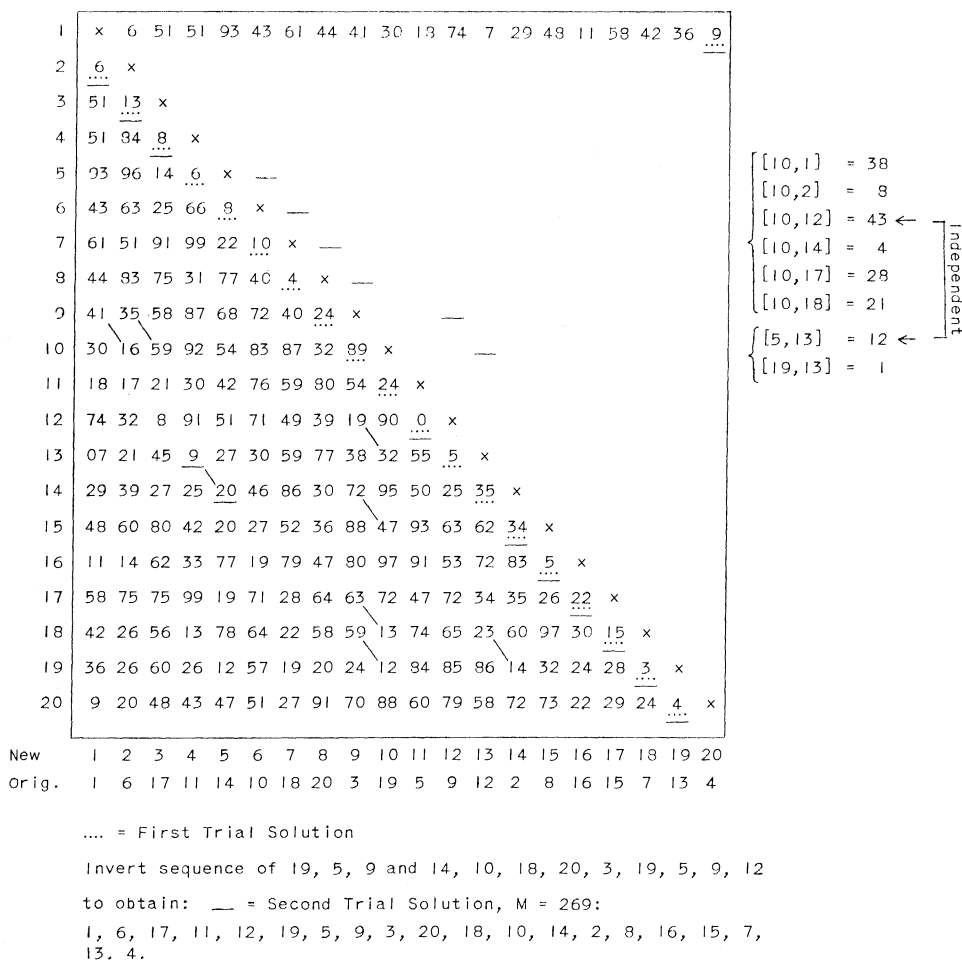


Figure 2

ments of the old trial solution, together with the new elements form a single cyclic sequence.

These two conditions will be used as a criterion to reject possible transformations. For rejection of a transformation it is sufficient to ascertain that it does not obey either of the two. A priori rules that ensure ad-

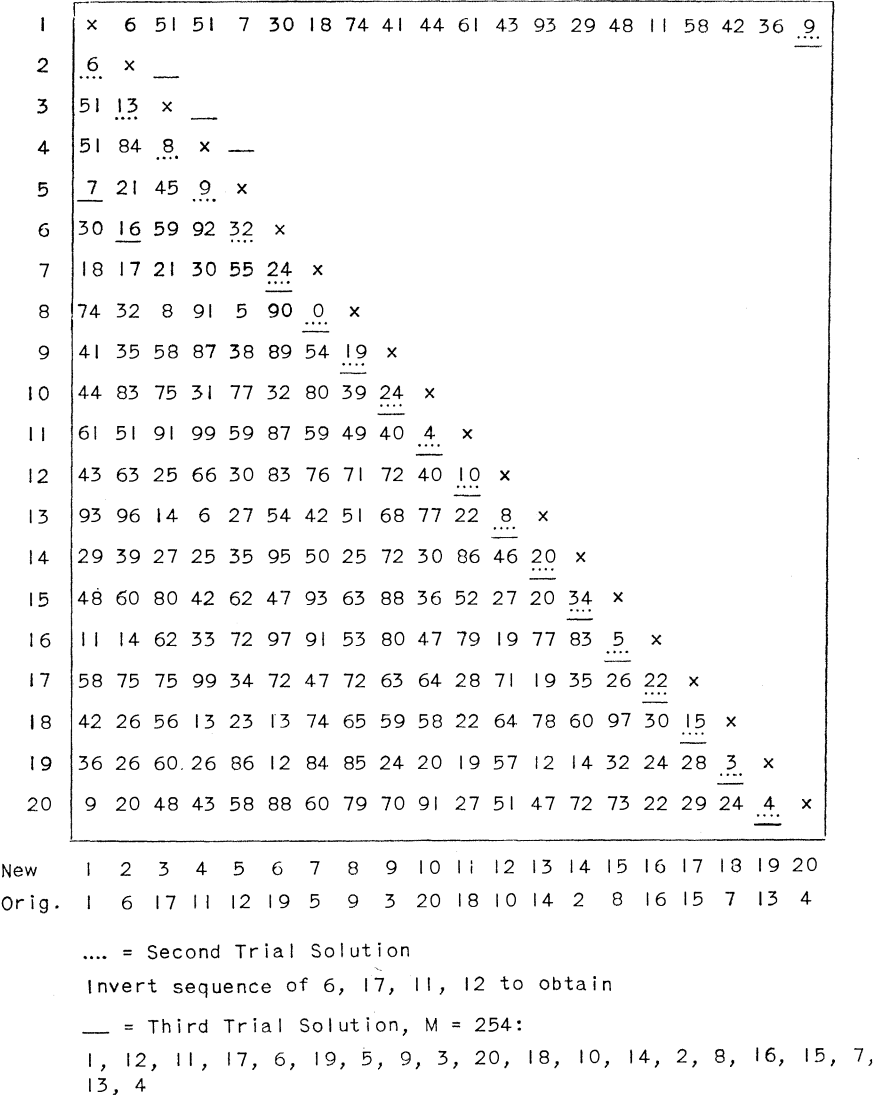


Figure 3

herence to condition (b) are difficult to handle in practice. Consequently condition (a) will be mainly used as a criterion.

Before the method is described, a way will be given to simplify the matrix and thus the subsequent discussions. This simplification is made possible by the following theorem, that has already been mentioned else-

As the last two terms of the right hand part of this equality are independent of x , $\sum_i \sum_j x_{ij} d_{ij}$ is minimal if and only if $\sum_i \sum_j x_{ij} c_{ij}$ is minimal, from which the theorem follows.

The simplification then consists of making the elements of the trial solution zero by subtracting and/or adding appropriate amounts to the rows and/or columns of the matrix. The resulting matrix will have positive, negative, and zero elements; the elements of its slant are all zero. From the theorem it is seen that this matrix has the same solution as the original one.

Condition (a) for the admissibility of a transformation now reduces to $\sum_N d_{ij} < 0$. For the sake of convenience we shall split the sum up into two parts, the sum over the positive elements included in the transformation $\sum_{N^+} d_{ij}$ and that over the negative elements $\sum_{N^-} d_{ij}$. The previous condition may then be written $\sum_{N^+} d_{ij} < -\sum_{N^-} d_{ij}$. We call minus the sum of all the negative elements in the matrix its 'residual' R and observe that $R \geq -\sum_N d_{ij}$. This leads to a weaker form of condition (a): *Condition (a')* One or more elements are not simultaneously elements of the solution if the sum of their values exceeds the residual of the matrix.

From this formulation it is seen that, in order to eliminate the largest number of positive elements, it is advantageous to make R as small as possible. In the actual description of the method below, an algorithm will be given that achieves this goal automatically by means of a series of additions and subtractions.

The main use of condition (a') will be to eliminate certain elements from the solution. An important consideration here is that when we have eliminated an element as a possible element of the solution by applying condition (a') to the matrix D , then this elimination remains in force if we change D by adding or subtracting constants to its rows and columns, although in the resulting matrix this elimination cannot be motivated by an application of condition (a').

Condition (a') as stated is rather general and consequently difficult to apply. Our procedure starts out with more specialized versions of it. When all eliminations by means of the weakest version have been carried out, the next weakest version is applied, etc. By the time that strong and complicated conditions are required, the number of elements to be considered has been reduced to such an extent that the resulting computations do not entail an undue strain upon our capabilities.

The weakest form of condition (a') that we shall use is: (i, j) is not an element of the solution, if $d_{ij} \geq R$.

To derive the next condition we observe that if (i, j) is an element of the solution, then the condition that there can be only one element in every row and column ensures that the elements of the present trial solution that are in the same row (or column) as (i, j) [i.e., the elements $(i, i+1)$

and $(j-1, j)$ are not elements of the solution. As this leaves row $i+1$ and column $j-1$ without solution elements, we have to find new elements in this row (or column). Since the best that we can do is to take the smallest element out of those from which we can choose, we can write down the condition: (i, j) is not an element of the solution if $d_{ij} + \min_k d_{k, i+1} + \min_l d_{j-1, l} \geq R$. If either of the last two terms on the left-hand side of this inequality is negative, it should be deleted (or what amounts to the same, put equal to zero), because it has already been included in R .

In the light of what has been said above, it will not be difficult to find other and stronger conditions along the same lines. The application of these conditions will be dealt with in the description below.

Among the elements that might be included in the solution, the contra-slant elements are of special interest, as they admit a stronger condition without undue complications. Inclusion of such an element not only deletes the elements of the present trial solution that are in the same row and column, but also the element that is symmetrical to the contra-slant element. This necessitates introduction of new elements in the row and column corresponding to this last element. The condition for the rejection of a contra-slant element can consequently be stated as follows: The contra-slant element $(i+1, i)$ is not an element of the solution if $d_{i+1, i} + \min_k d_{i-1, k} + \min_l d_{i, l} + \min_m d_{m, i+1} + \min_n d_{n, i+2} \geq R$. Care should be taken to avoid counting some of the members of the left-hand side twice.

A similar consideration applies to negative elements of the contra-slant. The lumping of all negative elements together in R was based on the rather optimistic assumption that we shall be able to find a transformation that includes all these elements. If one of these negative elements is a contra-slant element, however, inclusion of this element $(i+1, i)$ in the solution entails the introduction of new elements in column i and row $i+1$. This gives a means of reducing the value of R , viz. the rule: The maximum contribution of the negative contra-slant element $(i+1, i)$ to the residual of the matrix is $-d_{i+1, i} + \min_k d_{i, k} + \min_l d_{l, i+1}$, the last two terms again only to be taken into account if they are positive.

These additional properties of contra-slant elements are especially useful in the case of symmetrical problems. The slant elements contain the trial solution and therefore most of the small elements of the matrix. The contra-slant elements are equal to the slant elements, so that many of them will be negative in the simplified form of the matrix, which tends to increase R disproportionately.

Consequently it is advantageous from a computational standpoint to divide the transformations that we are looking for into two types: those that do not contain a contra-slant element and those that contain at least one contra-slant element.

The more detailed description of the method below is exemplified with

the problem used in the previous section, starting with the transformations that do not contain a contra-slant element.

Transformations without contra-slant elements. It should be emphasized that in all operations of this section (e.g., looking for the smallest element in a row) we completely neglect the contra-slant elements.

We start out with the matrix with the trial solution in the slant and subtract constants from the rows and columns in order to make the slant elements zero. One way to do this, which was found rather effective, is to subtract from each element in column i the amount $\min_j c_{ij}(j \neq i-1)$, then subtract from each element in row j the amount that is left in $(j-1, j)$.

In our example (Fig. 4) the amounts to be subtracted have been tabulated under the columns and to the right of the rows. The elements that become negative owing to these subtractions have been underlined with a dotted line, and their contribution to the residual has been listed row by row in the last column to the right of the matrix. In Fig. 5 these subtractions have been carried out; all elements larger than the total residual of 35 have been replaced by a dot. According to the weakest rule stated above, these elements cannot be elements of the solution.

Our next step is to reduce the magnitude of the residual. This is achieved by the following simple algorithm: Subtract from row j the amount $\min_i d_{ij}$ if:

1. this amount is greater than zero;
2. column $i=j-1$ contains at least one element $d_{ik} \leq 0$.

Add the same amount to column $i=j-1$ (row j and column $j-1$ are called 'adjoint'). Repeat this procedure until all amounts to be added and subtracted are zero.

Apart from the above algorithm there exists a second possibility of reducing the magnitude of R , if the number of negative elements in a row or column is larger than the number of nonpositive elements in the adjoint column or row. In that case we add the smallest negative number to the row, or column, and subtract it from the adjoint column, or row. This procedure can be repeated a few times, but its effects on the magnitude of R are generally only minor.

In Fig. 5 the tabulations to the left and over the matrix give the successive additions and subtractions to be applied to columns and rows as they were found. The first column to the right and the third row under the matrix give their sum total. Only row 15 still contains two negative numbers; adding 1 to row 15 and subtracting 1 from column 14 solves this situation and reduces R by 1. The second column to the right of the matrix gives again the contributions to R of the elements in each row, which sum to $R=10$.

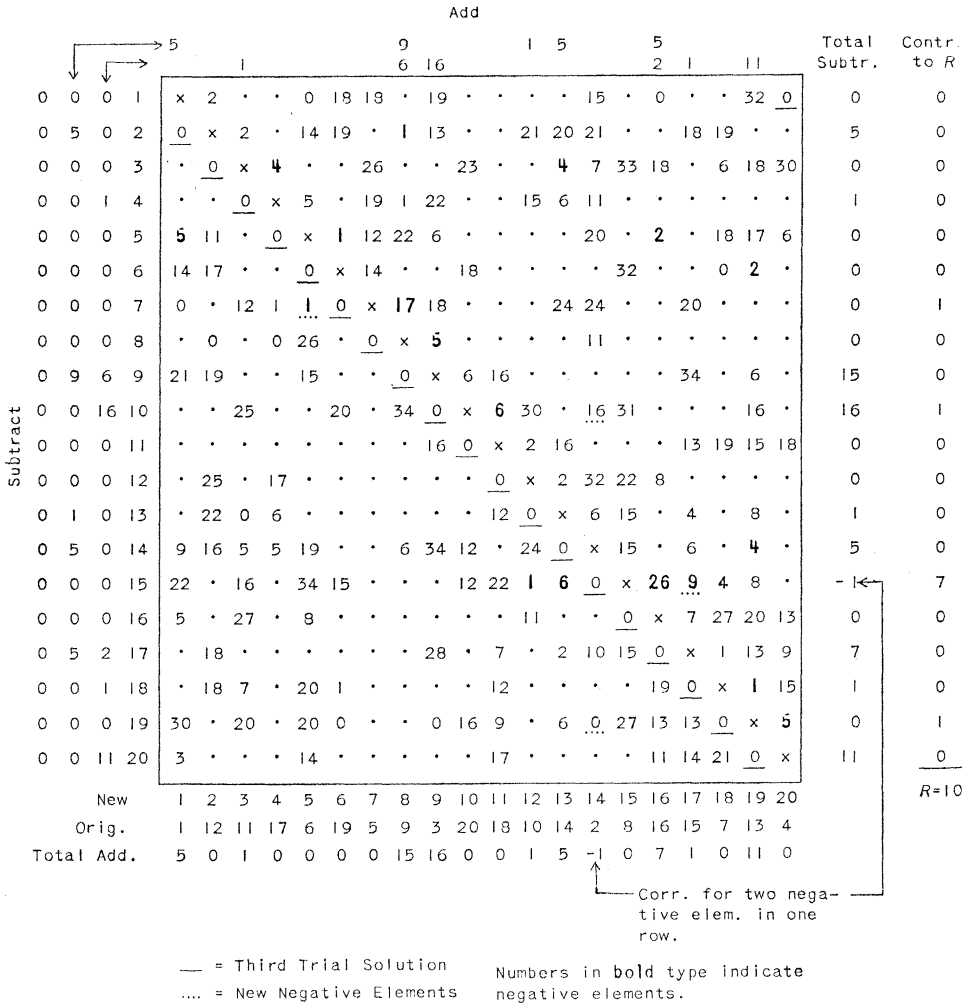


Figure 5

Figure 6 gives the new matrix, obtained by applying the additions and subtractions to the elements of matrix 5 and replacing all elements over 9 by a dot.

A further elimination of elements can be achieved by application of the next weakest form of the condition, i.e., the condition that (i,j) is not an element of the solution if $d_{ij} + \min_k d_{k,i+1} + \min_l d_{j-1,l} \geq R$. The elements that were eliminated by means of this rule are crossed out (X) in Fig. 6 [the elimination of (14,10) reduces R to 9].

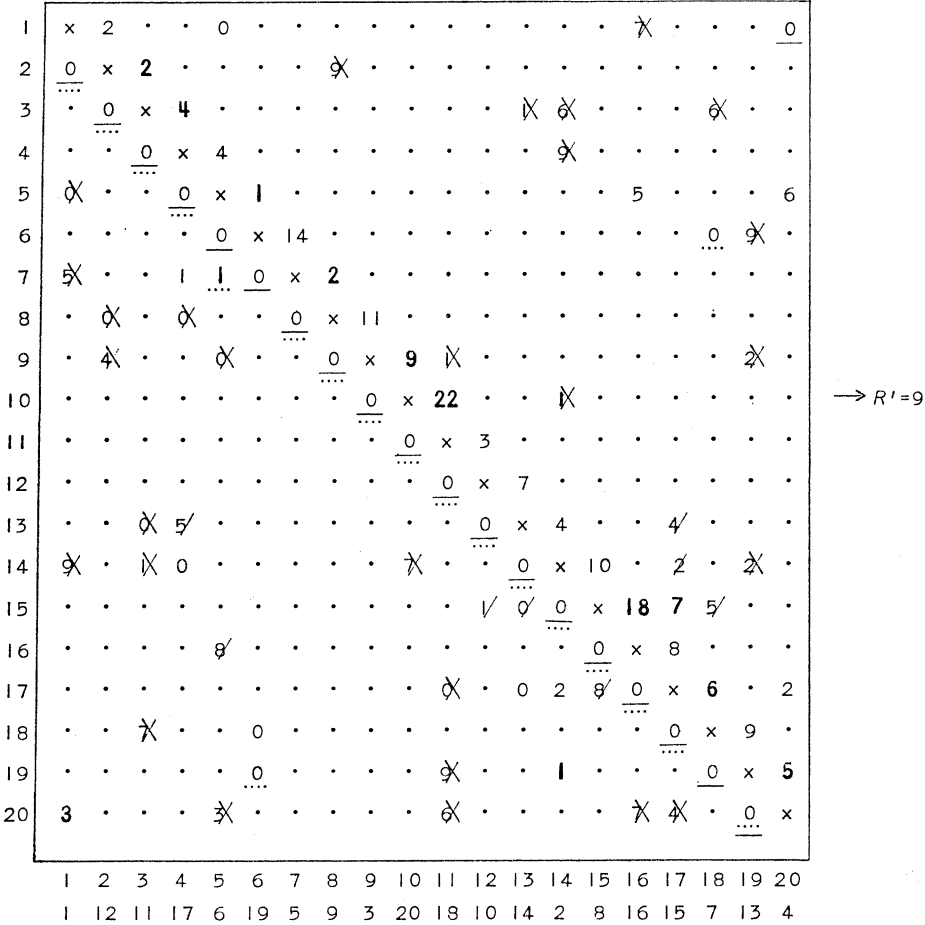


Figure 6

If a positive element is in the same row or column as a negative element (k,l) , the selection of this positive element as an element of the solution excludes the negative element from the solution. Consequently, the effective value of R , by means of which the eligibility for inclusion in the solution of such a positive element is judged, is $R+c_{k,l}$. The elements

eliminated by means of this rule and subsequent application of the previous one, have been crossed out by means of a single bar (/).

The elements of the matrix that are now still eligible for inclusion in the solution are few in number and we are in a position to investigate whether a legitimate transformation is possible. If the number of elements still had been large, we could have gone to the trouble of applying stronger conditions in order to eliminate more of them.

A way to construct all possible transformations out of the elements that are still eligible that was found quite convenient, is by means of 'trees.' A tree is constructed in the following way: Take any of the negative elements and write down next to it in a single column all the eligible elements in the row that is adjoint to the column of the negative element. Draw a bar from the first element (the 'first-generation' element) to every element in the column (the 'second-generation' element) and write alongside each bar the sum of the two elements that it connects. In an analogous way we find a family of third-generation elements for every second-generation element. Alongside the bar connecting a second- with a third-generation element we write the sum of the third-generation element and the amount alongside the bar connecting the first- and second-generation elements. This process can, in principle, be extended indefinitely.

The reasoning behind this construction is as follows: We want to find a transformation that includes the first-generation element. In order to include this element we have to delete the slant element in the same column from the trial solution. The second-generation elements are all possible elements that can fill the place of the deleted slant element in its row. Each second-generation element causes the deletion of another slant element and so brings about the birth of the third-generation elements, etc.

It is easily verified that, when in a branch an element is repeated, the insertion into the trial solution of the elements in this branch, from the repeated element up to its repetition to replace the appropriate slant elements, results in a 'solution,' such that $\sum_i x_{ij} = \sum_j x_{ij} = 1$. We still have to check whether this $\|x_{ij}\|$ also represents a single cyclic sequence and if so, whether it reduces the measure of the trial solution.

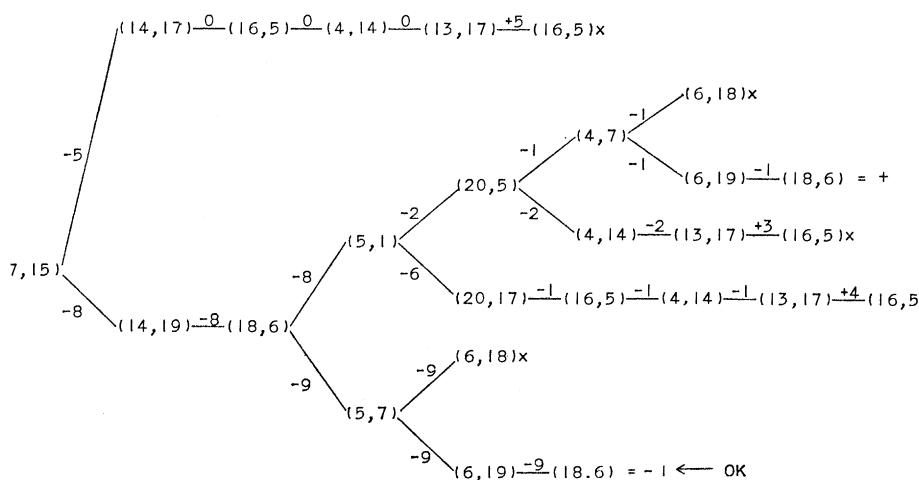
From the above it follows that we can stop extending a particular branch of the tree, as soon as one of its elements is repeated. The sum of its elements from the repeated element up to its repetition gives the amount by which the measure of the trial solution would be increased by the prospective transformation. If this amount is positive, we write behind the last element of the branch: $= +$; if it is negative, we write: $= -$.

Further it is seen that, as soon as the sum of all elements in a particular branch exceeds minus the sum of all negative elements that have not yet been included in that branch, any transformation that might result from

this branch cannot give a decrease in measure of the trial solution, so that we can terminate the branch. The first sum mentioned is given by the number written alongside the bar connecting the last element of the branch with the previous one. Termination of a branch because of this criterion is designated by the symbol: + behind the terminating element.

Finally, we know that a trial solution can never contain both the element (i,j) and the element (j,i) . Consequently, a branch can be terminated as soon as we reach the element (i,j) , if (j,i) is a previous element of the same branch. For this termination criterion we use the symbol: X.

In Fig. 7 the tree for matrix 6 that starts with the negative element (17,15) is given. After what has been said above, this figure should be



ly Possibility: (18,6) + (5,7) + (6,19)

Figure 7

self-explanatory. From it we see that the only desirable transformation (denoted by = -) is the one that replaces (5,6), (6,7), and (18,19) by (5,7), (6,19), and (18,6).

As a final test we have to check, whether this transformation results in a single cyclic sequence. This test is carried out following the procedure that we used to find the first trial solution, i.e., by means of construction of the new sequence. It shows that the above transformation results in the new sequence: 1,12,11,17,6,5,9,3,20,18,10,14,2,8,16,15,7, 19,13,4, whose measure is one unit smaller than that of the preceding solution.

At this point it is pertinent to note that a negative element has only

to be taken as the starting point of a new tree, if it does not occur in a previous tree. As all three negative elements of matrix 6 occur in the tree of Fig. 7, this tree alone is sufficient to deal with matrix 6.

The matrix, rearranged so as to get the new trial solution into its slant, has again to be checked for inversions and for transformations without contra-slant elements. The resulting computations have not been reproduced here, since they are of the same type as the previous ones. They show that no desirable transformations, of the type dealt with up to now, exist.

Transformations with at least one contra-slant element. The search for these transformations starts out in the same way as the preceding one. We take the matrix with the final trial solution in the slant and subtract or add numbers from rows and columns in the same way as previously, so that all slant elements become zero, and the residual, now taking into account the elements of the contra-slant, is as small as possible. This leads to the matrix given in Fig. 8, which has a total residual of 41, taking into account that the elements (13,19) and (20,19) cannot occur simultaneously in a transformation.

As mentioned above, the maximum contribution of a contra-slant element $(i, i-1)$ to the residual is

$$c_{i,i-1} + \min_k c_{i-1,k} + \min_l c_{l,i}. \quad (k \neq i, l \neq i-1)$$

In the column to the right of the matrix in Fig. 8 these maximum contributions have been tabulated row by row. They sum to $R=16$.

The deletion of positive elements on the basis of this smaller residual has now to be done very carefully, as these elements cannot be deleted if they contributed, by virtue of the above rule, to the reduction of the residual. This applies for instance to the elements (5,15) and (14,16). The elements that can be deleted have been crossed out (\times). For the sake of clarity the matrix with all the elements that are still eligible for inclusion in the solution is copied in Fig. 9.

It seems advisable, in this stage of the solution, to start with the construction of trees. However, due to the peculiar consequences of introducing contra-slant elements, a satisfactory way to present such trees is difficult to find. In order to be able to unearth the possibly present desirable transformations or to prove that such transformations do not exist, we shall change our technique slightly. Instead of assuming, as we did earlier, that there exists a desirable transformation that includes a specific negative element, we shall now assume that there exists a transformation that includes all negative elements (i.e., the elements that contribute to the residual) and investigate whether this assumption is tenable. This investigation is repeated for all possible different sets of one or more nega-

1	x	0	•	44	3	16	•	20	•	•	39	•	13	•	6	•	•	20	41	0	0
2	0	x	0	31	11	•	5	10	•	•	19	19	12	•	•	19	16	19	•	42	0
3	•	0	x	1	•	26	•	•	31	•	•	3	7	38	26	•	11	•	29	32	0
4	43	32	0	x	4	13	9	31	•	•	15	7	5	•	•	•	•	•	•	33	0
5	0	8	•	0	x	9	33	8	•	•	•	•	17	•	3	•	20	0	25	5	0
6	4	34	2	0	0	x	7	19	•	43	•	27	20	•	•	25	•	0	•	37	0
7	•	0	•	3	31	0	x	0	•	•	•	•	14	•	•	•	•	•	•	•	0
8	15	5	•	35	6	26	0	x	0	12	42	41	40	•	•	29	33	•	3	36	0
9	41	•	26	•	•	•	43	0	x	1	33	•	11	31	39	•	•	19	21	•	0
10	•	•	•	•	42	•	•	13	0	x	0	15	•	44	•	24	16	•	18	12	0
11	35	15	•	10	•	•	•	43	34	0	x	1	22	17	6	•	•	•	•	34	0
12	•	16	0	3	•	36	•	43	•	16	0	x	0	14	•	7	•	40	13	34	0
13	10	9	6	1	17	29	13	32	13	•	23	0	x	13	•	8	41	•	0	49	0
14	30	37	22	•	39	•	•	•	20	32	5	1	0	x	18	0	•	19	19	•	3
15	8	•	28	•	8	•	•	•	•	•	12	•	•	0	x	14	27	•	26	10	0
16	41	10	•	•	•	28	•	25	•	9	•	1	2	7	0	x	2	•	16	3	2
17	35	9	4	42	16	•	•	41	•	13	•	•	37	•	16	0	x	4	1	8	4
18	17	12	•	39	0	9	•	•	21	•	•	40	•	32	•	•	0	x	4	•	0
19	33	•	22	•	21	•	•	1	20	16	•	9	4	27	16	16	1	0	x	7	7
20	0	42	32	32	8	•	•	40	•	16	33	37	•	•	8	12	16	•	0	x	0
New	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	16
Old	1	12	11	17	6	5	9	3	20	18	10	14	2	8	16	15	7	19	13	4	

X' = Deleted because of introduction of *R'*.

Figure 8

tive elements. We find the solution of our problem by choosing from among all desirable transformations that one that gives the largest decrease in measure, and applying this transformation to the last trial solution. From the above we see that the procedure can be terminated as soon as we find a decrease in measure that is larger than minus the sum of the elements in any set of negative elements that we still have to investigate.

The investigation of the tenability of the assumption that a particular

1	x	0	•	•	3	•	•	•	•	•	•	•	13	•	6	•	•	•	•	<u>0</u>
2	<u>0</u>	x	0	•	11	•	5	10	•	•	•	•	12	•	•	•	•	•	•	•
3	•	<u>0</u>	x	1	•	•	•	•	•	•	•	•	3	7	•	•	•	11	•	•
4	•	•	<u>0</u>	x	4	13	9	•	•	•	•	15	7	5	•	•	•	•	•	•
5	0	8	•	<u>0</u>	x	9	•	8	•	•	•	•	•	•	3	•	•	0	•	5
6	4	•	2	0	<u>0</u>	x	7	•	•	•	•	•	•	•	•	•	•	<u>0</u>	•	•
7	•	0	•	3	•	<u>0</u>	x	0	•	•	•	•	•	•	•	•	•	•	•	•
8	•	5	•	•	6	•	<u>0</u>	x	0	12	•	•	•	•	•	•	•	•	3	•
9	•	•	•	•	•	•	<u>0</u>	x	1	•	•	11	•	•	•	•	•	•	•	•
10	•	•	•	•	•	•	13	<u>0</u>	x	0	15	•	•	•	•	•	•	•	•	12
11	•	•	•	10	•	•	•	•	<u>0</u>	x	1	•	17	6	•	•	•	•	•	•
12	•	•	0	3	•	•	•	•	•	<u>0</u>	x	0	14	•	7	•	•	13	•	•
13	10	9	6	1	•	•	13	•	13	•	<u>0</u>	x	13	•	8	•	•	0	•	•
14	•	•	•	•	•	•	•	•	•	5	1	<u>0</u>	x	18	0	•	•	•	•	•
15	8	•	•	•	8	•	•	•	•	12	•	•	<u>0</u>	x	•	•	•	•	10	•
16	•	10	•	•	•	•	•	•	•	9	•	1	2	<u>7</u>	<u>0</u>	x	2	•	•	3
17	•	9	4	•	•	•	•	•	•	•	•	•	•	•	•	<u>0</u>	x	4	1	•
18	•	12	•	•	0	9	•	•	•	•	•	•	•	•	•	<u>0</u>	x	4	•	•
19	•	•	•	•	•	•	•	1	•	•	•	9	<u>4</u>	•	•	•	1	<u>0</u>	x	7
20	0	•	•	•	8	•	•	•	•	•	•	•	•	8	12	•	•	<u>0</u>	x	•
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

.... = Trial Solution
 — = Final Solution.

Figure 9

set of first-generation elements gives a desirable transformation is carried out as follows:

1. Write down all the negative elements of the set under consideration.
2. Write down the complete set of smallest second-generation elements.
3. Check whether any of the second-generation elements is in the same row or column as a first-generation element. If it is, replace this second-generation element by the next smallest element in the column (row).
4. Check whether there are two or more second-generation elements in the same

row or column. If there are, replace that one by the next smallest element that causes the smallest increase in measure. Negative elements that were excluded from the first-generation elements are always ineligible.

5. Sum all elements. If this sum is positive, reject the hypothesis. If it is negative add the third-generation elements and repeat the above procedure, etc.

If we find a set for which all next-generation elements are already in the set and for which the sum of the elements is negative, we are blessed with a prospective transformation. As a final test we have to check in the customary way, whether this transformation results in a single cyclic sequence.

In our example we start with the set of first-generation elements: [(15,14), (17,16), (18,17), (20,19)]. The second-generation elements are: (13,12), (5,15), (14,13), (15,12), (1,20), (19,18), (2,1), which sum to +1. This rules out the possibility that the above four first-generation elements can occur simultaneously in the solution.

Analogous reasons rule out the simultaneous occurrence of most of the sets of first-generation elements. Prospective transformations were seldom found. An example is the set of first-generation elements [(13,19), (17,16), (18,17)], which results in a transformation that introduces as new elements: (13,19), (17,16), (18,17), (16,14), (15,5), (5,18), (4,6). This transformation, however, breaks the cyclic sequence up into two subsequences.

The most desirable transformation was found to be the one generated from the set [(15,14), (13,9)] bringing in the new elements (15,14), (13,19), (18,6), (5,15), and (14,16). It results in the new cyclic sequence: 1,12,11,17,6,16,8,15,7,19,5,9,3,20,18,10,14,2,13,4, that has a measure of 246. This sequence is the solution of the problem.

Final Remarks

The method presented in the foregoing has been tested on a large number of problems. The most prominent of these was Dantzig's 42-city problem,^[1] which was solved by hand in approximately 70 hours. It was a rather frustrating business, for after about 10 hours' work we had already found Dantzig's solution by means of inversions alone.

The only available datum about the speed of Dantzig's method is Heller's statement^[2] that a 9×9 symmetrical problem with random elements took on the average 3 hours to solve. Though such a small problem can hardly be considered a test for the present procedure, the calculation was carried through for a few of them with an average solving time of 25 minutes per problem.

The present method has the following advantages over Dantzig's method:

1. It seems to be appreciably faster.
2. The calculations can be terminated as soon as we deem the solution obtained so far sufficiently accurate for our purpose.
3. It can be mechanized to any degree desired (*see* below).
4. It can be adapted to solve asymmetrical problems (*see* final section).

As for the third point we note that the procedure does not contain subjective decisions and can thus in principle be programmed for a computer that has a sufficient storage capacity. The programming of stages 1 and 2 and that of the first step of stage 3 is straightforward. Programming of the rest will present some difficulties. Though these are certainly not insurmountable, it remains dubious whether it is efficient to use an electronic computer for these calculations, as they involve mostly inspectional work.

For those who work by hand, a number of labor-saving devices are available. They will find that most of their time is used for copying matrices in which they had to rearrange rows and columns. A fast and easy way of doing this is to use an IBM reproducing punch coupled to an IBM printer. Another method is to cut the columns of the matrix apart, arrange them in the new sequence, copy this, cut the rows apart, arrange them in the new sequence and copy again. The solving times mentioned above were determined the honest way: all copy work was done by hand.

Though the method appears to be better than anything else known for this problem, it is not yet entirely satisfactory and the culprit, of course, is stage 3: the final adjustment. There are a few possibilities to improve its efficiency and elegance, among them being the use of the symmetry of the original matrix and the introduction of a convenient dichotomy of transformations into those which transform a trial solution $1, 2, \dots, n$ into a single cyclic sequence and those which do not. Lack of time has prevented pursuit of this line of thought any further.

ASYMMETRICAL PROBLEMS

THE SYMMETRY of the initial matrix was used twice in the preceding developments:

1. In determining the decrease in measure of the trial solution by the inversion $I(i_1, i_k)$. For asymmetrical problems this decrease is given by

$$\Delta M(i_1, i_k) = \Delta M_{\text{sym}} + \sum_{i=i_1}^{k-1} (c_{i, i+1} - c_{i+1, i}).$$

The second term in the right-hand part of this equality can be taken care of easily by checking on the desirability of the inversions by sets for which i_1 is constant and i_k increases one unit at a time.

2. In determining the number of inversions that we have to check for desirability. For asymmetrical problems we have to check on all possible inversions.

Though from the above it is seen that asymmetrical problems will present more work, but hardly any difficulties, I should like to make one concluding remark on the application of the final adjustment procedure to asymmetrical problems. The two major reasons for carrying out the final adjustment for symmetrical problems in two steps were the high incidence of negative contra-slant elements and the drastic effect of introducing a contra-slant element into a new trial solution. Though for asymmetrical problems the first argument is in general not valid, the second argument was generally sufficient to warrant keeping the steps separate.

REFERENCES

1. G. DANTZIG, R. FULKERSON, AND S. JOHNSON, *Opns. Res.* **2**, 393-410 (1954).
2. I. HELLER, *Proceedings, Second Symposium on Linear Programming*, p. 643, Washington, 1955.
3. M. M. FLOOD, *Opns. Res.* **4**, 61-75 (1956).