The Method of Reduced Matrices for a General Transportation Problem*

PAUL S. DWYER AND BERNARD A. GALLER

University of Michigan, Ann Arbor, Mich.

In the usual statement of the transportation problem [1], one has N items to send from shipping points $A_1, \cdots, A_p$ to receiving points $B_1, \cdots, B_q$. The amounts available for shipment from $A_1, \cdots, A_p$ are $a_1, \cdots, a_p$, respectively, and the amounts needed at $B_1, \cdots, B_q$ are $b_1, \cdots, b_q$, respectively, where

$$\sum_{i=1}^{p} a_i = \sum_{j=1}^{q} b_j = N.$$

Given the cost c_{ij} of shipping one item from A_i to B_j, the problem is to allocate the shipments so as to minimize the total cost. Formally, the problem is to find a set of integers x_{ij} such that

$$x_{ij} \geqq 0, \tag{1}$$

$$\sum_{j=1}^{q} x_{ij} = a_i, \tag{2}$$

$$\sum_{i=1}^{p} x_{ij} = b_j, \tag{3}$$

$$T = \sum_{i=1}^{p} \sum_{j=1}^{q} c_{ij} x_{ij} \text{ is minimized.} \tag{4}$$

(Here x_{ij} is the number of items to be shipped from A_i to B_j, and T is the total cost.)

It could happen, however, that the shipments from A_i to B_j would have to pass through intermediate assembly points D_h. In this case a shipment from A_i to B_j would involve the cost c_{ijh} of shipping the item via assembly point D_h, and one would also have the capacities $d_1, \cdots, d_r$ of $D_1, \cdots, D_r$ to consider. Now the statement of the problem becomes:

Find a set of integers x_{ijh} such that

$$x_{ijh} \geqq 0, \tag{5}$$

$$\sum_{j=1}^{q} \sum_{h=1}^{r} x_{ijh} = a_i, \tag{6}$$

* Presented at the meeting of the Association, Aug. 27–29, 1956. This report was supported in whole or in part by the United States Air Force under Contract No. AF 41-(657)-9 monitored by the Operator Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Texas. This is not an official publication under the contract. Views or opinions expressed or implied herein are not to be construed as necessarily reflecting the views or indorsement of the Air Force or of the Air Research and Development Command.

$$\sum_{i=1}^{p} \sum_{h=1}^{r} x_{ijh} = b_j, \tag{7}$$

$$\sum_{i=1}^{p} \sum_{j=1}^{q} x_{ijh} = d_h, \tag{8}$$

$$T = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{h=1}^{r} c_{ijh} x_{ijh} \text{ is minimized.} \tag{9}$$

(Here x_{ijh} is the number of items to be shipped from A_i to B_j via D_h, and T is again the total cost.) In this way, one is led to consider a three-dimensional transportation problem and, more generally, if there are $k - 2$ intermediate sets of points, one is led to a k-dimensional problem. The problem as outlined here is only one of many applications of a more general linear programming problem. Another example is the group assembly problem [3], which requires the maximization of T.

The method most commonly used for the two-dimensional problem (i.e., as originally stated above) is the "simplex method" [2]. In this method one first finds a "feasible solution," i.e., a set of integers satisfying (1)–(3) above, and then successively improves the solution until (4) is satisfied as well. Two major difficulties with this method arise when one passes to the k-dimensional problem. The initial stage of finding a feasible solution becomes much more complicated than in the two-dimensional case, and the number of constraint equations grows so rapidly that the problem becomes unwieldy.

The purpose of this paper is to describe methods which avoid the first of these difficulties and lessen the second. A good feasible solution will be shown to be available in any number of dimensions, and the method of reduced matrices can be used to replace the simplex method and thereby even avoid the need for a preliminary feasible solution.

The method of reduced matrices is used to obtain the exact solution to the problem as stated above. It has been programmed to handle small problems with $k \leqq 7$ on the IBM 650, somewhat larger problems with $k = 2$ on the MIDAC at the University of Michigan, and very large problems with $k \leqq 20$ on the IBM 704 at the General Motors Technical Center in Detroit, Michigan.

In the description of the method below, the notation is that of the case $k = 3$, but in every case the statement for a general k is an immediate generalization of the statement which appears here.

The method of reduced matrices is based on the fact that a constant may be subtracted from each element of any row (column, etc.) of the cost matrix without changing the positions in the matrix where the final allocations should be made. Now a necessary condition for the existence of a solution to the problem is the existence of constants u_i, v_j, w_h such that

$$c_{ijh} - u_i - v_j - w_h = 0 \text{ whenever } x_{ijh} \neq 0, \tag{10}$$

and

$$c_{ijh} - u_i - v_j - w_h \geqq 0 \text{ in any case.} \tag{11}$$

If one subtracts the minimal element of each row (column, etc.) from each element of that row (column, etc.), (11) is preserved and one obtains a "reduced matrix" in which each row (column, etc.) contains at least one zero element, and (10) is satisfied at these matrix locations. (In the maximization problem, one changes the sign of all the c_{ijh} before starting the reduction. Then (11) is preserved after the first step, and the remainder of the reduction is the same as in the present case.) The zero elements thus obtained furnish a set of matrix positions for a possible allocation of items. Moreover, each subtraction increases the value of S, where

$$S = \sum_{i=1}^{p} u_i a_i + \sum_{j=1}^{q} v_j b_j + \sum_{h=1}^{r} w_h d_h . \tag{12}$$

It is clear from (6)–(8) and (11) that

$$\begin{aligned} S &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r} u_i x_{ijh} + \sum_{j=1}^{q}\sum_{i=1}^{p}\sum_{h=1}^{r} v_j x_{ijh} + \sum_{h=1}^{r}\sum_{i=1}^{p}\sum_{j=1}^{q} w_h x_{ijh} \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r} (u_i + v_j + w_h) x_{ijh} \leqq \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{h=1}^{r} c_{ijh} x_{ijh} = T, \end{aligned} \tag{13}$$

and the inequality becomes an equality when values of u_i, v_j, and w_h are determined so that (6), (7), (8), (10) and (11) are true. Then we have

$$S = T. \tag{14}$$

We may hence consider the problem as the dual one of the maximization of S, rather than the minimization of T, in (13).

One may attempt to find the solution by putting $x_{ijh} = 0$ whenever $c_{ijh} > 0$, so as to maximize S, but this commonly leads to a set of equations (6)–(8) in $x_{ijh} \neq 0$ which are inconsistent, i.e., there is no solution at all. It is possible then to use the information gained in the solution process to determine another transformation on the matrix which produces another zero in the resulting reduced cost matrix without losing any already at hand. This transformation leads to an increase in S, and after a finite number of such steps (usually quite small), one has $S = T$. At this point one has a solution to (6)–(8) that also satisfies (10) and (11) but not necessarily (5), as one or more values of x_{ijh} may be negative. Though negative solutions do not always appear, they do frequently result from this type of algebraic solution. They must be removed in order to arrive at an acceptable (positive) solution. This is accomplished by successive applications of additional transformations which eliminate these negative elements from the solution and are accompanied by an increase in S, although the zero term associated with the negative x_{ijh} becomes non-zero in the next reduced matrix. In this way all negative solutions are eliminated so that the conditions (5), (10), (11), and (14) are all satisfied.

At this stage we can guarantee that all x_{ijh} are non-negative, but we cannot guarantee that they are integral and hence they may not provide acceptable answers to the problem. We do know that they are rational, i.e., they are fractions or integers, since they are the solutions of simultaneous linear equations with integral coefficients.

We next note an important fact which is significant not only for this method of reduced matrices but also for any other method such as the simplex method which explicitly or implicitly uses (5), (10), (11), (13), (14), and rational x. This is that the basic relation

$$\max S = \min T \tag{15}$$

is attained with positive rational x_{ijh}. In this case there is no way by which the cost matrix may be reduced further, so as to provide an increment to S, without violating (10) or (11). In terms of the concepts of the method of reduced matrices, this means that no further reduction in a cost matrix is possible if the $x_{ijh} \geqq 0$ terms associated with the zero elements of the transformed matrix satisfy (6), (7) and (8), even though the values of the terms $x_{ijh} > 0$ are not integral.

We define a reduced matrix with integral $x_{ijh} > 0$ associated with its zero terms satisfying (6), (7) and (8) to be completely reduced since (15) is satisfied by positive integral values of x_{ijh} and no increment to S is possible by further reduction. A similar reduced matrix with associated $x_{ijh} > 0$ fractional is said to be finally reduced even though the integral solution with minimal sum is not identifiable from its zero terms. The important point is that every finally reduced matrix is a completely reduced matrix for a related problem with the same cost matrix but with frequencies

$$a_i' = La_i , \qquad b_j' = Lb_j , \qquad d_h' = Ld_h . \tag{16}$$

Here L is the least common multiple of all the denominators of the x_{ijh} so that the solution of the related problem is integral. Thus every finally reduced matrix, being also a completely reduced matrix of a related problem, cannot be further reduced. So the maximization of S may lead to an integral solution or to a fractional solution. In the latter case no further reduction of the cost matrix is possible.

The determination of an integral solution from a fractional solution is thus, in a very real sense, a new combinatorial problem. An additional element or elements must be selected to supplement those used in obtaining the fractional solution. The finally reduced matrix is much more satisfactory than the original cost matrix for this purpose. If a pure combinatorial approach is to be used, it is effective to take $x_{ijh} = 1$ corresponding to the smallest non-zero value of the finally reduced matrix and to adjust the x_{ijh} values of the fractional solution so as to satisfy (6)–(8). If this does not produce an integral solution, the next largest element of the finally reduced matrix is used. Commonly, in practical problems with unique solutions and appreciable differences in the c_{ijh} terms, an integral solution results very soon.

Though useful for handling many practical problems, this pure combinatorial method is not theoretically satisfactory. To provide a more satisfactory theoretical solution, a method has been derived which utilizes the fractional solution and the relations of the reduction process in determining the optimal integral solution. In a sense the fractional solution serves as a particular solution and the additional element plays the role of a parameter. The equations resulting from the previous reduction process provide necessary conditions for the integral solution which enable one to eliminate from consideration all but a subset of the elements of the finally reduced matrix. Under certain conditions this subset is relatively very small [5].

The integral solution then results from the addition to the fractional solution of the part contributed by the parameter. In this sense, the solution of this second combinatorial problem is integrated with the maximization of S, even though it is not possible to make further formal reduction of the cost matrix once a rational solution has been identified.

The determination of the best integral solution using one additional element has been programmed for the IBM 704.

A detailed presentation of the method of reduced matrices as adapted to machine computation is given in another paper [6].

A feasible, or approximate, solution considered here is based on the calculation of the weighted deviates of the cost matrix (c_{ijh}). If we write

$$(17) \qquad c_{i*h} = \sum_{j=1}^{q} b_j c_{ijh}, \qquad c_{**h} = \sum_{i=1}^{p} \sum_{j=1}^{q} a_i b_j c_{ijh},$$

and so forth, so that the asterisk represents weighted summation over the subscript whose position it is in, the formula for the two-dimensional weighted deviates is

$$(18) \qquad c'_{ij} = c_{ij} - \frac{c_{i*}}{N} - \frac{c_{*j}}{N} + \frac{c_{**}}{N^2},$$

while the three-dimensional formula for the weighted deviates is

$$(19) \qquad c'_{ijh} = c_{ijh} - \frac{c_{i**}}{N^2} - \frac{c_{*j*}}{N^2} - \frac{c_{**h}}{N^2} + \frac{2c_{***}}{N^3}.$$

Similar formulas are available for the k-dimensional problem.

If one uses the matrix of deviates instead of the original cost matrix, one can obtain a good approximate solution to the original problem. More exactly, one obtains a set of integers x_{ijh} satisfying (5)–(8) and which are close to satisfying (9). The method consists of choosing the minimal weighted deviate (i.e., algebraically minimal) and allocating as many items as possible to its position in the matrix, then finding the minimal weighted deviate among the remaining ones and allocating as many of the remaining items to its position as possible, etc. As indicated below, this method furnishes an efficient approximate solution in an extremely short time. Especially in problems in which the original data are subject to error, this approximate solution may sometimes be used instead of the true minimal solution. Moreover, if such an approximate solution were used as a first feasible solution in the simplex method, it should reduce considerably the number of iterations necessary to reach the optimum cost.

Programs for the calculation of the weighted deviates and the approximate solution have been written for the IBM 650 ($k \leqq 7$) and the IBM 704 ($k \leqq 20$). Typical of the time required for an approximate solution is two minutes, thirty-five seconds for a problem with $k = 4$, $N = 137$, $p_1 = p_2 = p_3 = p_4 = 3$ on the 650, while the same problem took only twelve seconds on the 704. (We now write $p_1, p_2, \cdots, p_k$ instead of $p, q, r, \cdots$.) A large problem ($k = 2$, $N = 3077$, $p_1 = 186$, $p_2 = 15$) ran 19.5 minutes on the 704, producing a solution which is 98 percent efficient. Efficiency is defined by the formula:

$$E = \frac{\bar{T} - T_a}{\bar{T} - T} \tag{20}$$

where T_a is the cost resulting from the approximate solution, T is the true minimal cost, and $\bar{T} = c_{\cdots}/N^2$ is the mean cost, which is the cost one would obtain if every cost element were replaced by the weighted mean of all cost elements. It should also be mentioned that the same weighted deviates are used for the maximization form of the problem (such as in the group assembly problem) as well as in the minimization form which occurs in the transportation problem.

REFERENCES

[1] HITCHCOCK, F. L., The distribution of a product from several sources to numerous localities. *J. Math. Physics*, *20*, 1941, 224–230.

[2] DANTZIG, G. B., Application of the simplex method to a transportation problem. Chap. XXIII, Activity analysis of production and allocation. Cowles Commission Monograph No. 13, New York, Wiley, 1951.

[3] ROBY, T. B., Problems of rational group assembly exemplified in the medium bomber crew. San Antonio, Texas: Human Resources Research Center, Lackland Air Force Base, July, 1953. (Research Bulletin 53–18).

[4] The theory described here is abstracted from a report, "Maximum group assembly sums," and other reports prepared by Paul S. Dwyer, for the Director, Detachment 4 (Crew Research Laboratory), Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas.

[5] DWYER, P. S. *A Theorem on the Multi-Index Transportation Problem*. Abstract. *Bull. Am. Math. Soc.* *63* (1957), p. 14.

[6] GALLER, B. A., AND DWYER, P. S. *Translating the Method of Reduced Matrices to Machines*. *Naval Research Logistics Quarterly* *4* (1957), pp. 55–71.