SIMPLIFIED TREATMENT OF DEGENERACY IN TRANSPORTATION PROBLEMS*

BY

KURT EISEMANN

International Business Machines Corporation, New York

The solution of transportation problems by the methods described in Refs. [1, 2, 3] may give rise to degenerate distributions which have less than $m + n - 1$ non-zero elements in a matrix of size $m \times n$. Inasmuch as a complete basis of $m + n - 1$ entries is required at each stage, two methods have been proposed for handling degenerate cases (Refs. [1, 3]). Both of these methods have undesirable drawbacks. It is the purpose of this paper to present a procedure which combines the advantages of both methods without any of their disadvantages. The procedure involves a negligible amount of work and is equally well suited for incorporation into high-speed computer codes as well as for hand computations.

Where degeneracy arises in the middle of an iteration by the presence of more than one basis element as admissible candidate for elimination, only one entry is liquidated, all other competing elements remaining as explicit basis entries of value zero. Once a proper basis has been obtained at the beginning of the problem, the number of basis elements (some of them possibly with value zero) present during all iterations will remain unaltered and degeneracy to fewer basis elements can no longer arise. Cycling is obviated by the application of Charnes' perturbation method, which, when applied to the transportation problem, takes the following form**:

If several x_{ij} are tied for removal from the basis, choose only x_{IJ} by the criterion $I = \min$ (tied i). If ties remain, choose $J = \min$ (tied j).

This can also be formulated thus: If $x_{i_\mu j_\nu}$ ($\mu, \nu = 1, 2, \cdots$) are tied, choose (i_M, j_N) such that $ni_M + j_N = \min (ni_\mu + j_\nu)$.

Or, shorter, always choose the northwesternmost element for elimination.

When this rule is applied throughout, cycling cannot occur.

The difficulty then remains of converting a degenerate initial distribution into a proper basis. The ϵ-technique proposed by Dantzig [1] resolves this problem. On the other hand, the method proposed by Charnes [3] amounts to augmenting the basis by the insertion of strategically distributed zero entries. The choice of locations for these basis elements is by no means arbitrary. Many unfilled positions in the matrix are inadmissible for this purpose as they would not provide a proper basis. Even among the admissible locations, which permit some degree of choice, distribution of zeros must take place judiciously. With large matrices this may pose a formidable problem. A systematic method of accomplishing a valid allocation of zero elements is provided by the previously mentioned ϵ-technique: after the initial distribution is completed, all table entries are rounded off. Multiples of ϵ are thereby converted to zeros, which now occupy admissible positions, and other basis elements are thereby restored to the values they would bear if the epsilons had never been introduced in the first place.

However, for application to an automatic computer, the ϵ-technique introduces undesirable features: either the word size of data must be curtailed to accommodate the adjoined multiples of ϵ — not an exhilarating prospect; or the multiples of ϵ must be carried as separate words during the initial distribution, involving auxiliary book-keeping as well as the tying up of further storage space, usually dictating a reduction in the maximum size matrix that the machine can conveniently handle. The procedure described below accomplishes the proper allocation of zeros without any of these objectionable features.

Many schemes can be devised for an initial distribution that provides a "basis" and at the same time strives to reduce the number of necessary iterations. The "north-west corner rule" [3] takes no account of the entries in the cost matrix and therefore provides a distribution which is basic, but which usually requires, for large-scale problems, an exorbitant number of iterations. A better scheme would be the selection of the lowest cost element, allocation of the maximum permissible quantity to this route, and con-tinuation by considering cost elements of successively increasing values. Alternatively, one might take into account cost coefficients as well as their first differences in rows and columns, together with pertinent quantities. For any such method, examples can be constructed for which a different method would give a distribution closer to the optimum than the method proposed.

As a compromise, one of the most convenient and reasonably efficient methods for machine applications is the one incorporated into existing high-speed computer codes. Let D_i = quantity required at destination i, S_j = quantity available at source j, c_{ij} = unit cost for the route from j to i, x_{ij} = quantity to be allocated to route (i, j), "filled" column or row = a column or row whose entire quantity has been allocated ($S_j = 0$ or $D_i = 0$), "unfilled" column or row = a column or row still possessing a disposable quantity ($S_j \neq 0$ or $D_i \neq 0$).

Consider the first row, $i = I = 1$.

1. Select the lowest cost element in this row, $c_{IJ} = \min_j c_{Ij}$

2a. If the selected column is unfilled ($S_J \neq 0$), allocate the maximum possible quantity to this route, $x_{IJ} = \min (D_I , S_J)$, replacing D_I and S_J by $D_I - x_{IJ}$ and $S_J - x_{IJ}$

2b. If the selected column has previously been filled ($S_J = 0$), make no allocation to column J but proceed with step 3.

3. If the current row remains unfilled, select the next lowest cost element in the same row and repeat step 2, until this row has been completely filled.

4. Proceed to the next row and repeat steps 1, 2, 3.

In this way only one row of the cost matrix need be available in random-access storage at any one time and the matrix need be read-in from auxiliary storage only once in normal row order.

Imagine the basis entries connected by horizontal and vertical lines, joining the nearest elements within the same column or within the same row. Degeneracy can then be shown to lead to a disconnected graph: the distribution falls into two (or more) disjoined independent sections, for each of which the graph consists of an unbroken chain of interconnected links. For each section, the sum of the row totals, $\sum D_i$, equals the sum of the column totals, $\sum S_j$. Placing zeros strategically then means the de-termination of those positions, for which the additional horizontal and/or vertical lines introduced by the zeros will result in an interconnection of the separated chains.

Consider the example of the diagram below. x denotes the allocated basis elements. Neighboring entries have been joined by horizontal and vertical lines, resulting in three disconnected chains, labeled for convenience as subscripts to the x's. A numbered entry such as "23" means that allocation of a zero basis element to the particular route will result in an interconnection of chains Nos. 2 and 3.

1 2	x_1	13	12	13	11	12
x_2	12	23	x_2	23	12	x_2
12	x_1	13	12	13	x_1	12
23	13	x_3	23	x_3	13	23
22	12	23	x_2	23	12	22
23	13	33	23	x_3	13	23

It is now clear that an admissible basis, which requires a single connected graph for the entire matrix, dictates the following choice of locations for additional zero entries:

Out of the three groups of locations labeled 12, 13, 23, choose any one location for each of two groups out of the three. No entry may be made for locations labeled 11, 22, and 33.

All this may be achieved automatically by a device of extreme simplicity.

Whenever an allocation is made under step 2a above, either a column or a row turns from "unfilled" to "filled". Degeneracy, i.e. disconnectivity of the graph, is now seen to result in the situation that whenever the last allocation to one of the independent chains is made, *both* the affected row *and* column quantities become exhausted *simultaneously*, $D_I \rightarrow 0$ as well as $S_J \rightarrow 0$. Recognition of this fact points to an immediate remedy: start with all columns and rows "unfilled". Then combine steps 1–4 enumerated above with the following:

5. Whenever the quantity of a column or row becomes exhausted, $S_J \rightarrow 0$ or $D_I \rightarrow 0$, mark it as "filled". One such marking must be made for each allocation in the table.

6. Whenever the quantities of column and row become exhausted simultaneously, $S_J \rightarrow 0$ as well as $D_I \rightarrow 0$, mark only the *column* as "filled", leaving the row "unfilled".

This procedure has the effect that on the subsequent application of step 3, the machine will find the current row unfilled and will therefore proceed to make a further allocation of the remaining quantity $D_I = 0$ of this row. Strict application of step 2b then prescribes that this allocation must invariably be made to unfilled columns, never to filled ones. Connectivity of the current chain with subsequent chains is thereby ensured.

Defective bases, the phenomenon of degeneracy, are thus painlessly rehabilitated.

It should be noted that other methods for obtaining an initial distribution may be modified by an analogous procedure. In particular, arbitrarily specified degenerate initial distributions, e.g. to represent current operations, may be made basic, if the

chains connecting the table entries contain no closed loops: run through the given allocations row by row, applying steps 5 and 6 above. As soon as a zero quantity remains in an unfilled row, allocate it to any unfilled column, preferably the one with lowest c_{ij}.

For machine applications, indication of filled columns is provided by a mere zero test of residual column quantities.

Example.

Quantities

22	F 2	2	F 3	7	8
1			1		
4	2		2		
5					
5					
7					

Costs

8	7	5	9	8
1	7	4	5	5
1	3	0	8	5
3	1	3	7	4
2	5	5	3	0

In the above table, distribution has been made for the first two rows. After allocation of element x_{21}, columns 1 and 3 have been marked as "filled", row 2 is still "unfilled" with the residual quantity 0. The next procedural step requires its allocation to one of the unfilled columns. Continuing, we obtain, after the allocation of row 4:

	F	F	F		F
22	2	2	3	7	8
1			1		
4	2		2		0
5		2			3
5					5
7					

The same situation now results in the allocation of quantity 0 to the unfilled column 4. The complete basic initial distribution becomes

22	2	2	3	7	8
1			1		
4	2		2		0
5		2			3
5				0	5
7				7	

Note that if the last zero had been allocated as element x_{41}, x_{42}, or x_{43} in violation

of our rules, the resulting distribution would not have been basic, thwarting application of the iterative methods.

The method described herein is currently being incorporated into the code now in preparation for the IBM.705 high-speed electronic computer.

References

1. G. B. Dantzig, *Application of the simplex method to a transportation problem*, in *Activity analysis of production and allocation*, J. Wiley and Sons, 1951, p. 359
2. A. Henderson and R. Schlaifer, *Mathematical programming, Appendix*, Harvard Business Review **32**, 94 (1954)
3. A. Charnes and W. W. Cooper, *The stepping stone method of explaining linear programming calculations in transportation problems*, Management Science 1, 49 (1954)