

Systems of Distinct Representatives and Linear Algebra*

Jack Edmonds

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(November 16, 1966)

Some purposes of this paper are: (1) To take seriously the term, "term rank." (2) To make an issue of not "rearranging rows and columns" by not "arranging" them in the first place. (3) To promote the numerical use of Cramer's rule. (4) To illustrate that the relevance of "number of steps" to "amount of work" depends on the amount of work in a step. (5) To call attention to the computational aspect of SDR's, an aspect where the subject differs from being an instance of familiar linear algebra. (6) To describe an SDR instance of a theory on extremal combinatorics that uses linear algebra in very different ways than does totally unimodular theory. (The preceding paper, Optimum Branchings, describes another instance of that theory.)

Key Words: Algorithms, combinatorics, indeterminates, linear algebra, matroids, systems of distinct representatives, term rank.

1. Introduction

The well-known concept of term rank [5, 6],¹ is shown here to be a special case of linear-algebra rank. This observation is used to provide a simple linear-algebra proof of the well-known SDR theorem. Except for familiar linear algebra, the paper is self-contained.

Incidentally to SDR's, an algorithm is presented for computing the determinant or the rank of any matrix over any integral domain. It is a variation of Gaussian (i.e., linear) elimination which has certain advantages. It is observed to be an interestingly bad algorithm for computing term rank.

The final part of the paper discusses some simple matroidal aspects of SDR's.

2. Systems of Distinct Representatives

Let Q be any finite family of subsets of a finite set E . "Different" members of Q may be identical in content. The number of members of a family H is denoted by $|H|$. The union of the members of H is denoted by $\cup(H)$. The SDR theorem says that it is possible to choose a different element from each member of Q if and only if there is no subfamily H of Q such that $|H| > |\cup(H)|$. The "only if" part is obvious.

A subset of E formed by choosing a different element from each member of Q is called a *system of distinct representatives* of Q , or an *SDR* of Q . (Sometimes, such as in [1], it is called a transversal of Q .

However, here the word "transversal" will be used differently.)

3. Matrices of Zeros and Ones

The subject of SDR's is frequently treated in the context of matrices of 0's and 1's. The *incidence matrix* of the family Q of subsets of E is the matrix $A = [a_{ij}]$, $i \in E$, $j \in Q$, such that $a_{ij} = 1$ if $i \in j$, and $a_{ij} = 0$ otherwise.

A *matching* in a matrix is a subset of its positions (i, j) such that first indices (rows) of members are all different and second indices (columns) of members are all different. A *transversal* (*column transversal*) of a matrix is a matching in the matrix which has a member in each column. The product of the entries of a transversal, we call a *transversal product*.

Clearly, Q has an SDR if and only if its incidence matrix A has a 1-valued transversal (or a 1-valued transversal product).

In the context of 0,1 matrices the SDR theorem says that a matrix $A = [a_{ij}]$, $i \in E$, $j \in Q$, of 0's and 1's has a 1-valued transversal if and only if there is no $H \subseteq Q$ and $K \subseteq E$ such that $|H| > |K|$ and such that submatrix $[a_{ij}]$, $i \in K$, $j \in H$, contains all the 1's of submatrix $[a_{ij}]$, $i \in E$, $j \in H$.

4. Permanents, Determinants, and Good Algorithms

The maximum cardinality of a nonzero-valued matching (matching having no zero entries) in any matrix A is well-known as the *term rank* $r(A)$ of A .

*Prepared while the author was a visiting professor at the University of Waterloo, Ontario.
¹ Figures in brackets indicate the literature at the end of this paper.

Clearly $r(A)$ is the maximum cardinality of a subset of columns of A which has a nonzero-valued transversal. Thus the *term rank* of a family of sets is defined to be the maximum cardinality of a subfamily which has an SDR.

Clearly $r(A)$ equals the maximum order of a square submatrix of A which has a nonzero-valued transversal. The *permanent* of a square matrix is defined to be the sum of all its transversal products. Thus $r(A)$, for a nonnegative matrix A , equals the maximum order of a submatrix of A with non-zero permanent. The permanent of a square 0,1 matrix is the number of 1-valued transversals of the matrix.

A good algorithm is not known for computing the permanent of any square 0,1 matrix (relative to the order of the matrix as the "size" of input). For an algorithm to be *good* we mean that there is a polynomial function $f(n)$ which for every n is an upper bound on the "amount of work" the algorithm does for any input of "size" n .

The transversals of any square matrix partition uniquely into two classes such that any two transversals which differ by just two positions (of each) are in different classes. (The proof is left to the reader.) The *determinant* of a square matrix with a prescribed 1-1 correspondence between rows and columns, i.e., with a prescribed transversal, is defined to be the sum of transversal products over whichever of these classes contains the prescribed transversal minus the sum of transversal products over the other class.

A very remarkable formal property of determinants, of say integers, is that there exists a good algorithm for computing them — a version of Gaussian elimination which we will describe in section 7. It is also remarkable that there exists a good algorithm, the same one as above, for computing the linear-algebra rank of any matrix of say integers. The largest number of digits of an entry as well as the dimensions of the matrix must of course be figured somehow into the measure of "size" of possible inputs to the algorithm. For example, the "size" of an input may be taken to be the maximum of these numbers, or may be taken to be a vector consisting of these several numbers.

An algorithm which is good in the sense used here is not necessarily very good from a practical viewpoint. However, the good-versus-not-good dichotomy is useful. It is easily formalized (say, relative to a Turing machine, or relative to a typical digital computer with an unlimited supply of tape), and usually it is easily recognized informally. Within limitations it does have practical value, and it does admit refinements to "how good" and "how bad." The classes of problems which are respectively known and not known to have good algorithms are very interesting theoretically.

Good algorithms, not Gaussian elimination, are well-known for computing the term rank of any 0,1 matrix. For instance see a remark in section 8 of this paper together with [3]. We shall see that Gaussian elimination is also an algorithm for computing term rank.

5. Rank

A main point of this paper is the following observation. Tutte in [7] uses the same idea in a deeper way.

THEOREM 1. *The term rank of a 0,1 matrix A is the same as the linear-algebra rank of the matrix obtained by replacing the 1's in A by distinct indeterminates (over any integral domain).*

Several "numbers" are called indeterminates if no polynomial function of them (over the given integral domain) equals zero unless it is identically zero.

PROOF OF THEOREM 1. The linear-algebra rank of a matrix equals the maximum order of a minor with nonzero determinant. (A *minor* of a matrix is a submatrix with a prescribed 1-1 correspondence between its rows and columns.) The determinant of a minor is a linear combination, with +1's and -1's coefficients, of all its transversal products. Thus, where the entries are zeros and distinct indeterminates, a determinant is nonzero if and only if one or more of its transversal products is nonzero. End of proof.

6. A Linear-Algebra Proof of the SDR Theorem

Suppose a matrix A of zeros and distinct indeterminates has no transversal of nonzeros. Then, by Theorem 1, its rank is less than the number of columns, so the columns are linearly dependent. Let A_1 be a submatrix consisting of a minimal dependent subset H of the columns of A . Let A_2 be a submatrix consisting of a maximal independent subset K of rows of A_1 . Since row rank equals column rank, we have $|H| = |K| + 1$.

If A has a column of all zeros then that column alone is a dependent set H , and K is empty. Otherwise, there is a column vector x such that $A_1 x$ is all zeros and such that the components of x are nonzero polynomials in the entries of A_2 .

If any row, a , of A_1 , not in A_2 , contained one or more indeterminates, then ax would be a polynomial function of indeterminates, equal to zero, but not identically zero. Hence, the rows of A_1 , not in A_2 , contain only zeros. End of proof.

7. A Bad Algorithm for Term Rank

Theorem 1 also suggests using Gaussian elimination as an algorithm for computing term rank. With the excuse of trying to give this approach every possible advantage, we describe an improved version of Gaussian elimination (possibly due to Gauss).

The rank of any matrix A^0 , or the determinant of any matrix A^0 (with mutually corresponding index sets), with entries from an integral domain, can be computed as follows.

In each step, k , beginning with $k=1$, obtain a new matrix $A^k = [a_{ij}^k]$ from $A^{k-1} = [a_{ij}^{k-1}]$, by first choosing some row, call it $i(k)$, and some column, call it $j(k)$, neither of them chosen on previous steps and such that $a_{i(k),j(k)}^{k-1} \neq 0$. The algorithm stops when this is no longer possible.

The matrix A^k is determined as follows. Each row $i(h)$, $1 \leq h \leq k$, is the same in A^k as in A^{k-1} . Each column $j(h)$ of A^k is all zeros except for entries $a_{i(g)j(h)}^k$, $1 \leq g \leq h$. For any other row i and any other column j ,

$$a_{ij}^k = (a_{i(k), j(k)}^{k-1} a_{ij}^{k-1} - a_{i, j(k)}^{k-1} a_{i(k), j}^{k-1}) / a_{i(k-1), j(k-1)}^{k-1}$$

where $a_{i(0), j(0)}^0$ is defined to be 1. It follows from Theorem 2 that this quotient exists in the integral domain (without fractionalizing).

THEOREM 2. *The a_{ij}^k , as defined above, is equal to the determinant of the minor of A^0 whose rows and columns are as listed in the brackets:*

$$\begin{pmatrix} i(1) & i(2) & \dots & i(k) & i \\ j(1) & j(2) & \dots & j(k) & j \end{pmatrix}$$

The rank of matrix A^0 equals the number of steps in the algorithm.

PROOF: Since each row $i(h)$, $1 \leq h \leq k$, is the same in A^k as in A^{k-1} , we may denote

$$a_{i(h)j(h)}^{h-1} = a_{i(h)j(h)}^h = a_{i(h)j(h)}^{h+1} = \dots = a_{i(h)j(h)}^k$$

simply by $a_{i(h)j(h)}^*$, and denote $a_{i(0)j(0)}^0 = 1$ by $a_{i(0)j(0)}^*$. Each of the other rows, i , of A^k is obtained from elementary row operations in A^{k-1} in a manner such that the determinant of any minor of A^k , containing all of the rows and columns, $i(h)$ and $j(h)$, and say m other rows and m other columns, is different from the determinant of the corresponding minor in A^{k-1} by a factor of $(a_{i(k)j(k)}^* / a_{i(k-1)j(k-1)}^*)^m$. (Recall from the theory of determinants that replacing any single row by c times that row multiplies the value of the determinant by c , and that subtracting any multiple of one row from another row does not change the value of the determinant.)

By iterating the above relationship between determinants, for the minors

$$\begin{pmatrix} i(1) & i(2) & \dots & i(k) & i \\ j(1) & j(2) & \dots & j(k) & j \end{pmatrix}$$

of A^0 , A^1 , \dots , and A^k , we find the determinant of this minor of A^k is different from the determinant of this minor of A^0 by a factor of

$$\begin{aligned} & \left(\frac{a_{i(1)j(1)}^*}{a_{i(0)j(0)}^*} \right)^k \left(\frac{a_{i(2)j(2)}^*}{a_{i(1)j(1)}^*} \right)^{k-1} \dots \left(\frac{a_{i(k)j(k)}^*}{a_{i(k-1)j(k-1)}^*} \right)^1 \\ & = a_{i(1)j(1)}^* a_{i(2)j(2)}^* \dots a_{i(k)j(k)}^* \end{aligned}$$

On the other hand, this same minor of A^k has determinant equal to

$$a_{i(1)j(1)}^* a_{i(2)j(2)}^* \dots a_{i(k)j(k)}^* a_{ij}^k,$$

because it is triangular $a_{i(k)j(h)} = 0$ for $1 \leq h < g \leq k$, and $a_{ij(h)}^k = 0$ for $1 \leq h \leq k$, and it has that value as the

product of its diagonal elements. Therefore the determinant of the corresponding minor of A^0 is equal to a_{ij}^k .

The statement that the rank of A^0 equals the number of steps of the algorithm follows in the same way as the same statement for a usual form of Gaussian elimination (the one where you divide so as to get 1's in a diagonal, or the one where you cross multiply without dividing): Where n is the total number of steps, the final A^n has the same rank as A^0 since it is obtained from A^0 by elementary row operations (which are reversible). The rank of A^n is n because A^n contains the minor

$$\begin{pmatrix} i(1) & i(2) & \dots & i(n) \\ i(1) & i(2) & \dots & j(n) \end{pmatrix},$$

having nonzero determinant, and because any row of A^n not included in this minor is all zeroes. End of proof.

Crucial implicit parts of our algorithm are the algorithms for arithmetic operations in the appropriate integral domain. These include an algorithm for dividing one member of the domain by another whenever that quotient exists in the domain.

The algorithm for matrices A^0 of integers is good. This follows from (a), (b), and Theorem 2. (a) The familiar algorithms, for arithmetic operations on integers in arabic notation, are good relative to the numbers of digits in the operands. (b) A determinant of order n is a sum of $n!$ terms, each, except for sign, being a product of n entries. Where m is the largest number of digits in an entry, nm is an upper bound on the number of digits in one such term, and so $(nm + \log n!) < n(m + \log n) < n(m + n)$ is an upper bound on the number of digits in the value of the determinant. (The base of the log is the base of the arabic notation.) Theorem 2 tells us that each entry of each A^k is the determinant of a minor of A^0 . Thus, relative to the number of rows of A^0 , the number of columns of A^0 , and the maximum number of digits in any entry of A^0 , we get an algebraic upper bound on the work of carrying out the computation of each a_{ij}^k . Since, relative to the number of rows of A^0 and the number of columns of A^0 , there is an algebraic upper bound on the number of such computations in the algorithm, we get an algebraic upper bound on the amount of work in the whole algorithm.

The present algorithm is more efficient, for say an integer-valued A^0 , than either the fractional or the purely cross-multiplicational versions of Gaussian elimination, because it divides out common integer factors of the entries, typically the only common integer factors, thereby keeping down the number of digits in entries, and thus the amount of arithmetic work. Of course, most computational uses of Gaussian elimination are approximating procedures where the entries are repeatedly rounded off to a smaller number of digits. When used with the same extent of round-off, the present algorithm is no less work than any other version of Gaussian elimination—in fact, more work because of all the divisions to do. However, I suspect that, when used with the same extent of round-off, it is more accurate.

Curiously, the algorithm for matrices of zeros and distinct indeterminates is not good. For such matrices, the a_{ij}^k 's are polynomials in indeterminates. Algorithms for arithmetic operations on polynomials in indeterminates are evident; these algorithms are good relative to "size" of the polynomials. However, using a known way to represent the polynomials, their "size" in general increases too fast relative to k . In particular, a_{ij}^k might have as many as $k!$ terms.

A problem suggested by these considerations is to find a good algorithm for computing the rank of any matrix whose entries are polynomials (or monomials), with integer coefficients, in any number of indeterminates not necessarily different in different entries. Presumably such an algorithm, if one exists, would somehow combine Gaussian elimination with the techniques for term rank. It would be interesting to determine whether there exists a good algorithm for computing the rank of any matrix whose entries are zeros and indeterminates not necessarily distinct.

8. Transversal Matroids

A *matroid* $M=(E, F)$ is a set E of elements and a nonempty family F of so-called independent subsets of E such that (1) every subset of an independent set is independent, and (2) for any $S \subseteq E$, all maximal independent subsets of S contain the same number of elements, called the rank $r(S)$ of S . (F contains at least the empty set.) Where independent subset of columns means linearly independent subset of columns, the set of columns in a matrix (over any integral domain) forms a matroid, called *the* matroid of the matrix.

In [1] it was shown by an elementary combinatorial argument that for any finite family Q of sets, the subfamilies of Q which have SDR's are the independent sets of a matroid whose elements are the members of Q . Such matroids were dubbed *transversal matroids*.

It is not clear a priori that, from a good algorithm for deciding whether any given family Q has an SDR, we can obtain a good algorithm for finding the term rank of any Q , i.e., the maximum cardinality of a subfamily of Q which has an SDR. To take an analogy, a good algorithm is known for deciding whether the members of any given Q are mutually disjoint, but a good algorithm is not known for finding, for any Q , the maximum cardinality of a subfamily with this property.

The fact that any Q forms a matroid, as described, says that we can find the term rank of any Q as follows. Consider the n members of Q one after another in any predetermined order. Let L^0 be the empty subfamily of Q . At the k th step ($k=1, \dots, n$) determine whether the family consisting of $L^{k-1} \cup Q$ together with the k th member of Q has an SDR. If it does, then it is L^k . Otherwise, $L^k=L^{k-1}$. The cardinality of L^n is the term rank of Q .

Suppose, given any Q with a nonnegative numerical weight corresponding to each member, that we wish to find a maximum weight-sum subfamily of Q which has an SDR. It is proved in [2] that we get such a subfamily as L^n of the preceding algorithm simply

by first ordering the members of Q according to non-increasing weights. Some other consequences of the matroid property of SDR's are developed in [1].

Can every transversal matroid be represented as the matroid of a matrix? By appealing to Theorem 1, we see that the answer is yes: the matroid whose independent sets are the subfamilies of Q having SDR's is the matroid of the matrix obtained from the incidence matrix A of Q by replacing the ones by distinct indeterminates. Indeed, Theorem 1, together with the well-known fact that matrices yield matroids as described, provides an immediate "nonelementary" proof that any finite family Q yields a matroid as described: Relative to the incidence matrix A of the family Q , the independent sets of the transversal matroid of Q are the subsets of columns that have 1-valued transversals. Theorem 1 says that, upon replacing the 1's by distinct indeterminates these are precisely the linearly independent subsets of columns.

(The matroid of the incidence matrix A is not generally a transversal matroid. In fact the matroid of A generally depends on what integral domain its zeros and ones come from. One can of course define the transversal matroid of any matrix, so that relative to it the rank of a subset of columns equals term rank of those columns.)

Notice by considering the transpose of incidence matrix A , that the rows of A are the elements of a matroid whose independent sets are the subsets of rows that have 1-valued "row transversals." In other words, for any family Q of subsets of a finite set E , the members of E are the elements of a matroid for which the independent sets are the subsets of E that are SDR's of subfamilies of Q . A corresponding version of the SDR theorem says that a subset $P \subseteq E$ is an SDR of some subfamily of Q if and only if there is no $S \subseteq P$ that has nonempty intersection with fewer than $|S|$ members of Q .

A number of papers have developed a certain other relationship between SDR's and linear algebra which (along with much else on the subject) we do not discuss here. This was one of the purposes of Hoffman and Kuhn in [4]. Their other purpose was to extend the SDR theorem to characterize when a family Q of subsets of E has an SDR which contains a prescribed subset $P \subseteq E$. In view of the SDR theorem, their conditions are (1) that Q have an SDR, and (2) that P be an SDR of some subfamily of Q .

As a final illustration here of the present linear-algebra viewpoint, we prove the Hoffman-Kuhn theorem: If set $P \subseteq E$ is an SDR of some subfamily of Q , then, by the matroid property for such sets, P can be extended to some maximum cardinality subset of E , say T_0 , which is an SDR of some subfamily of Q , say Q_0 . If family Q itself has an SDR, say T , contained in E , then we have $|Q|=|T|=|T_0|=|Q_0|$, and thus $Q=Q_0$. End of proof.

This paper was originally submitted to the American Mathematical Monthly. I am indebted to the referee for several changes in wording and for the following comments:

"The well-known SDR theorem" is not well-known to the Monthly readers and the author should give credit at some point to Philip Hall.

Using integers ij etc., to represent both sets and elements is logically inconsistent. Thus " i " represents both an element of E and a set of Q . The author should stick to conventional means.

The author is unnecessarily sophisticated in defining "minor of a matrix." It can be simply and clearly defined as a square submatrix.

The author should give a reference for a conventional proof of Philip Hall's SDR theorem, e.g., Halmos and Vaughn (Amer. J. Math. **72**, 1950) which I believe to be at least as simple as the proof given here.

9. References

[1] Jack Edmonds and D. R. Fulkerson, Transversals and matroid partition J. Res. NBS **69B** (Math. and Math. Phys.) No. 3, 147-153 (1965).

- [2] Jack Edmonds, Matroids and the greedy algorithm, in preparation.
[3] Marshall Hall, Jr., An algorithm for distinct representatives, Amer. Math. Monthly **63**, 716-717 (1956).
[4] A. J. Hoffman and H. W. Kuhn, Systems of distinct representatives and linear programming, Amer. Math. Monthly **63**, 455-460 (1956).
[5] Oystein Ore, Theory of Graphs (Chapter 7), Amer. Math. Soc. Colloq. Publs, **38** (1962).
[6] Herbert John Ryser, Combinatorial Mathematics (Chapters 5 and 6), Carus Mathematical Monographs, No. 14, Math. Assoc. Amer. (1963).
[7] W. T. Tutte, The factorization of linear graphs, J. London Math. Soc. **22**, 107-111 (1947).

(Paper 71B4-250)