

Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics

Joseph B. Kruskal
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

A three-way array X (or three-dimensional matrix) is an array of numbers x_{ijk} subscripted by three indices. A triad is a multiplicative array, $x_{ijk} = a_i b_j c_k$. Analogous to the rank and the row rank of a matrix, we define $\text{rank}(X)$ to be the minimum number of triads whose sum is X , and $\dim_1(X)$ to be the dimensionality of the space of matrices generated by the 1-slabs of X . (Rank and \dim_1 may not be equal.) We prove several lower bounds on rank. For example, a special case of Theorem 1 is that

$$\text{rank}(X) \geq \dim_1(UX) + \text{rank}(XW) - \dim_1(UXW),$$

where U and W are matrices; this generalizes a matrix theorem of Frobenius. We define the triple product $[A, B, C]$ of three matrices to be the three-way array whose (i, j, k) element is given by $\sum_r a_{ir} b_{jr} c_{kr}$; in other words, the triple product is the sum of triads formed from the columns of A , B , and C . We prove several sufficient conditions for the factors of a triple product to be essentially unique. For example (see Theorem 4a), suppose $[A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$, and each of the matrices has R columns. Suppose every set of $\text{rank}(A)$ columns of A are independent, and similar conditions hold for B and C . Suppose $\text{rank}(A) + \text{rank}(B) + \text{rank}(C) \geq 2R + 2$. Then there exist diagonal matrices Λ , M , N and a permutation matrix P such that $\bar{A} = AP\Lambda$, $\bar{B} = BPM$, $\bar{C} = CPN$. Our results have applications to arithmetic complexity theory and to statistical models used in three-way multidimensional scaling.

I. INTRODUCTION

A three-way array (or three-dimensional matrix) is an array of numbers x_{ijk} , subscripted by three indices. We let $i = 1$ to I , $j = 1$ to J , $k = 1$ to K . A *triad* is a three-way array which has multiplicative form, that is, an array X for which $x_{ijk} = a_i b_j c_k$. We are interested in the decomposition of an array X

into triads,

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr}.$$

The minimum number R of triads which are needed is called the *rank* of X .

A ν -*slab* of X is a matrix formed by fixing the ν th index, for $\nu = 1, 2$, or 3 . The dimensionality of the linear space generated by the ν -slabs of X is called $\dim_\nu(X)$. Array rank subsumes matrix rank, and \dim_ν subsumes row rank and column rank in the following sense: if the I -by- J -by- K array degenerates to an I -by- J matrix because $K = 1$, then it is easy to see that array rank becomes matrix rank, $\dim_1(X)$ and $\dim_2(X)$ become row rank and column rank, and $\dim_3(X)$ is 1 (or 0). Similar results hold if $I = 1$ or $J = 1$ instead. Of course, $\text{rank}(X)$ and $\dim_\nu(X)$ are not usually equal, though it is true that $\text{rank}(X) \geq \dim_\nu(X)$.

Multiplication of a matrix by an array to give an array can be defined in a natural way (see Sec. 2 for details). We prove (see Theorem 1) that

$$\text{rank}(X) \geq \dim_1(UX) + \text{rank}(XW) - \dim_1(UXW).$$

This generalizes a matrix theorem of Frobenius (1911). We also prove (see Corollary 1 to Theorem 2) that

$$\text{rank}(X) \geq \dim_1(X) - 1 + \min \text{rank}\{uX : \text{all } u \text{ such that } uX \neq 0\}.$$

Here u is a vector, so that uX is a matrix formed by taking a linear combination of the 1-slabs of X .

Let A be the I -by- R matrix of elements a_{ir} , and similarly for B and C . The reverse of decomposition is provided by the *triple product* $[A, B, C]$ of three matrices, which we define to be the array whose (i, j, k) element is $\sum_r a_{ir} b_{jr} c_{kr}$. Note the analogy with the ordinary matrix product AB' (where B' means B transpose). A triple product can be taken only when all three matrices have the same number of columns. If they each have R columns, we shall say that the triple product *involves* R columns.

We prove some lower bounds on rank in terms of some given decomposition. For example (see Theorem 3a), suppose $X = [A, B, C]$ and the decomposition involves R columns. Write I_0 for $\text{rank}(A)$, J_0 for $\text{rank}(B)$, and K_0 for $\text{rank}(C)$. (Obviously $I_0 \leq R$, $J_0 \leq R$, $K_0 \leq R$.) Suppose *every* set of J_0 columns of B is independent, and *every* set of K_0 columns of C is independent.

Suppose that $J_0 + K_0 \geq R + 1$. Then

$$\text{rank}(X) \geq I_0 + \min(R - I_0, J_0 - 1, K_0 - 1, J_0 + K_0 - R - 1).$$

This lower bound lies between I_0 and R .

Suppose we have two different decompositions of the same array, so that $[A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$. What can we say about the relationship between the two decompositions? In other words, in what sense does the triple product determine its factors? It is easy to see that if P is a permutation matrix and Λ, M, N are diagonal matrices such that $\Lambda MN = \text{identity matrix}$, then $[A, B, C] = [AP\Lambda, BPM, CPN]$. The permutation matrix corresponds to a rearrangement of the triads, while the diagonal matrices cancel out, leaving each triad unchanged. We prove several sufficient conditions under which the decomposition is unique up to this kind of change, which we call *equivalence*. For example (see Theorem 4a), suppose $X = [A, B, C]$ is a decomposition involving R columns. Use I_0, J_0 , and K_0 for matrix ranks, as above. Suppose *every* set of I_0 columns of A is independent, and that similar conditions hold for B and C . Suppose $I_0 + J_0 + K_0 \geq 2R + 2$. Then any other decomposition of X involving R columns is equivalent to the given one.

A key lemma underlying this result is of interest in itself. To give the flavor of the Permutation Lemma, we state a special case. Suppose A and \bar{A} are two I -by- R matrices, and suppose A has no zero columns. Let $w(\text{vector}) = \text{the number of nonzero elements of the vector}$; suppose that for any vector x such that $w(xA) \leq R - \text{rank}(A) + 1$ we have $w(xA) \leq w(x\bar{A})$. Then there are a permutation matrix P and a nonsingular diagonal matrix Λ such that $A = AP\Lambda$. This lemma is reminiscent of matroid theory, but we have not been able to find it in the literature.

Application to Arithmetic Complexity

In the area of arithmetic complexity, several important operations can be described by a three-way array. For example, consider ordinary multiplication of two 2×2 matrices U and V to give a product W , $UV = W$. Index the elements of each matrix by a *single* index, i for U , j for V , and k for W , as shown in Fig. 1(a), where $i = 1$ to I , $j = 1$ to J , $k = 1$ to K , and in this case $I = J = K = 4$. Now every element w_k is a linear combination with fixed numerical coefficients of terms of the form $u_i v_j$. We let x_{ijk} be the coefficient of $u_i v_j$ in w_k ; in this case, each x_{ijk} is either 0 or 1, as shown in Fig. 1(b). More generally, whenever we wish to compute a set of bilinear forms, we can describe the desired results by using a three-way array of coefficients x_{ijk} :

$$w_k = \sum_{i=1}^I \sum_{j=1}^J x_{ijk} u_i v_j, \quad k = 1 \text{ to } K.$$

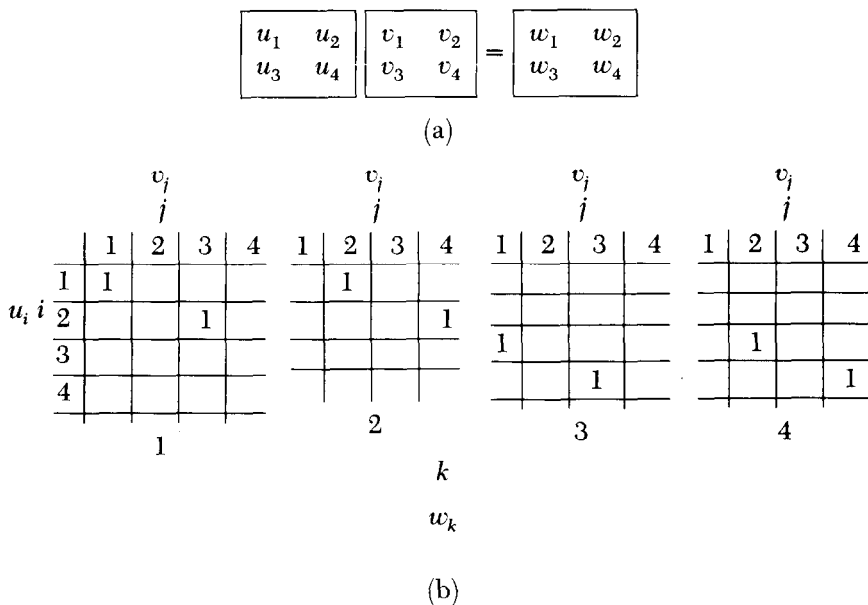


FIG. 1. (a) Multiplication of two 2×2 matrices. (b) Three-way array describing 2×2 matrix product.

Among other examples which fit this framework are multiplication of quaternions, multiplication of Cayley numbers, and multiplication of polynomials (modulo some fixed polynomial, if desired).

A decomposition of X corresponds to an algorithm for calculating the elements w_k . Thus suppose $[A, B, C] = X$. Then one method of calculating the w_k may be seen in the first expression:

$$\begin{aligned} \sum_r \left[\left(\sum_i a_{ir} u_i \right) \left(\sum_j b_{jr} v_j \right) c_{kr} \right] &= \sum_i \sum_j \left[\sum_r a_{ir} b_{jr} c_{kr} \right] u_i v_j \\ &= \sum_i \sum_j x_{ijk} u_i v_j = w_k. \end{aligned}$$

In words, this method is to calculate the $2R$ linear combinations $\sum a_{ir} u_i$ and $\sum b_{jr} v_j$, then multiply them pairwise to form R products, and then take the linear combinations of these products with coefficients c_{kr} .

		r							
		1	2	3	4	5	6	7	8
u_i	i	1	1		1				
		2		1		1			
		3					1		1
		4						1	1

v_j	j	1	1			1			
		2			1			1	
		3		1			1		
		4				1			1

w_k	k	1	1	1					
		2			1	1			
		3					1	1	
		4						1	1

(a)

		r						
		1	2	3	4	5	6	7
u_i	i	1	1	-1	1			1
		2		1			1	
		3				1		1
		4	-1			-1	1	-1

v_j	j	1	1			1	1		1
		2			1				1
		3					1	1	
		4	-1	1	1			1	

w_k	k	1	1	-1			1	1	
		2		1	1				
		3				1	1		
		4	1			1	-1		1

(b)

FIG. 2. (a) Ordinary method of matrix multiplication. (b) Strassen method of matrix multiplication

Figure 2 illustrates two different decompositions of the array X shown in Fig. 1(b). [Of course, both sets of three matrices yield the *same* triple product, shown in Fig. 1(b).] Figure 2(a) shows the decomposition which corresponds to the obvious natural method of matrix multiplication, while Fig. 2(b) shows the decomposition for Strassen's method [16]. The former requires $R=8$ triads, and hence 8 "active" multiplications (see below), while the latter requires only $R=7$.

A central concept in the field of arithmetic complexity is the "length" of a calculation, which refers to the minimum possible calculation time among all possible methods of calculation (in the sense of Strassen [18]). While a very general approach is sometimes taken to the time required for each elementary operation, in this particular context it is common to count each "active" multiplication as requiring unit time, and all other operations as requiring zero time. An active multiplication is one in which both factors involve the input variables. Not counting divisions is justified, since Strassen [19] has proved that if each division also requires unit time, then the use of divisions cannot reduce the length (for the type of calculation we are discussing). Not counting additions, subtractions, and inactive multiplications can be justified in several situations.

(a) In one important application the elements u_i and v_i are large matrices, and the active multiplications are much more expensive than the other operations.

(b) When asymptotic results are desired (as the size of the matrices or other objects gets large), it turns out in many situations that this simplification does not effect the exponent of growth, but only the constant multiplier.

(c) For many algorithms, the entries in A , B , and C turn out to be very simple values like ± 1 , so the inactive multiplications are much cheaper.

(d) On analog computers, an active multiplication requires a multiplication of two variables, which is far more expensive than the other operations.

Accordingly, we define

$\text{length}(X)$ = the minimum possible number of active multiplications in making the calculation described by X .

In the method of calculation based on a decomposition of X into R triads, there are exactly R active multiplications. Clearly $\text{length}(X) \leq \text{rank}(X)$. In order to state some precise results from complexity theory, we temporarily assume that the numbers we are dealing with need not be real numbers, but merely belong to a suitable ring. It is known that if the u_i and v_i are noncommuting indeterminates (or are isomorphic to such indeterminates), then $\text{length}(X) = \text{rank}(X)$; if the u_i and v_i are commuting indeterminates (or are isomorphic to them), then $\text{rank}(X)/2 \leq \text{length}(X) \leq \text{rank}(X)$. For proofs

of these results, see, e.g., [19] or [12]. Other related results will be cited in the next section, when an adequate notation is available.

Application to Statistics

“Canonical decomposition” [6] is a data analysis method which is based on finding the least-squares *approximate* decomposition, involving R columns, of a given three-way data array X . For our purposes, R should be considered as a fixed number which we are given. It is beyond the scope of this paper to describe how R is chosen in practice. We merely mention that the choice involves subjective human judgment, and that R is almost always too small to permit an exact decomposition. Effective methods now exist to determine the decomposition numerically. (We remark however that possible improvements in speed offer fertile territory for numerical analysts, especially by comparison with the heavily worked field of bilinear numerical analysis.)

The importance of canonical decomposition rests primarily on its use as a method of computation for another data analysis method called INDSCAL [6] which is the most successful variety of “individual differences multidimensional scaling”. An explanation of these data analysis methods and their interrelationships is outside the scope of this paper. We merely state that after the data have received some preliminary processing, canonical decomposition is applied. Since the resulting matrices have R columns, each row can be considered as a point in R -dimensional space. In statistical use, each matrix is considered as a configuration of points in R -dimensional space.

One major virtue of INDSCAL, which has contributed greatly to its success, is that the configuration of points is not freely rotatable, but has a fixed orientation with respect to the coordinate axes. By contrast, virtually no related methods have this property, including multidimensional scaling, factor analysis, principal components (in a certain sense), “points-of-view analysis”, three-mode factor analysis, and three-mode multidimensional scaling. This property rests on the same property for canonical decomposition. Because of the importance of this property, a full mathematical understanding of it is desirable. The sufficient conditions in this paper generalize those in [13], which may be consulted for further explanation and other references.

II. SOME LOWER BOUNDS ON RANK: THEOREMS 1 AND 2

The main purpose of this section is to present some lower bounds on the rank of a three-way array X . However, we also introduce other rank-like numbers for an array, and a variety of elementary facts and inequalities. We

also cite some results from complexity theory and show their relationship to our results. We suppose throughout that X is an I -by- J -by- K array, x_{ijk} .

The rank of an array is sensitive to the domain of numbers used in the decomposition. The same array can have different ranks over different domains, as Howell [12] explores in some detail, so it is important to specify the domain when stating results in this area. We shall assume that the domain consists of the real numbers in this paper, although many of our results can be generalized in varying degrees.

Although the arrays we deal with are not individually symmetric, many of the *concepts* we deal with are symmetric under permutation of the subscripts. Just as the rank of a matrix remains the same under transposition, $\text{rank}(X)$ remains the same under the five possible transpositions of X (based on the six permutations of three subscripts). However, $\text{length}(X)$ does *not* remain the same under all transpositions. Just as a matrix has the symmetric concepts of row rank and column rank (which may not be equal for matrices over a ring), an array has three rank-like numbers analogous to row rank and column rank. However, these three numbers, which are the dimensionalities of certain vector spaces, are generally not equal to one another nor to the rank, even though we consider only arrays of real numbers.

To avoid undue repetition and complex notation, we shall generally state definitions and results only in one form, and shall rely on the reader to supply the symmetric concepts and results. For example, we shall use X_i to indicate the i th slab of X , which is a J -by- K matrix; of course there are I such slabs. We shall not use any explicit notation for the slabs in other directions, though slabs in all directions are equally important to us. We define $\text{dim}_1(X)$ to be the dimensionality of the space consisting of all linear combinations of the X_i , and we define dim_2 and dim_3 similarly in terms of the slabs in other directions. These three numbers are analogous to row rank and column rank of a matrix.

A *representation* (A, B, C) is the same thing as a decomposition. We define the *rank of a representation* (A, B, C) to be the number of columns in each of the matrices A, B, C . We define the *rank of an array* X as the smallest rank of any representation of X by a triple product, $X = [A, B, C]$. Thus an array has four kinds of rank-like numbers: dim_1 , dim_2 , dim_3 , and rank. If the I -by- J -by- K array X degenerates to an I -by- J matrix (that is, if $K = 1$), then it is easy to see that the definition of $\text{dim}_1(X)$ specializes to the row rank of X , the definition of $\text{dim}_2(X)$ specializes to the column rank of X , the value of $\text{dim}_3(X)$ is 0 or 1, and the definition of $\text{rank}(X)$ as an *array* specializes to the definition of $\text{rank}(X)$ as a *matrix*. The last of these facts is particularly fortunate, for it means that in discussing rank we do not need to distinguish between array rank and matrix rank in cases where both would

be meaningful. It is easy to prove that

$$\dim_1(X) \leq \text{rank}(X).$$

Of course the same holds for \dim_2 and \dim_3 . It is also easy to see that

$$\max_i \text{rank}(X_i) \leq \text{rank}(X) \leq \sum_i \text{rank}(X_i),$$

and that

$$\text{rank}(X) \leq JK - \text{card}\{(j, k) | x_{ijk} = 0 \text{ for all } i\} \leq JK.$$

If $X = [A, B, C]$, then this representation provides some more elementary bounds, most obviously $\text{rank}(X) \leq \text{rank}(A, B, C)$. If the representation has rank R but can be reduced to one of lower rank, this leads to better bounds:

$$\text{rank}(X) \leq R - \text{number of zero columns of } A;$$

if r columns of A have rank 1 and the *corresponding* r columns of B have rank 1, then $\text{rank}(X) \leq R - r + 1$.

Now we introduce multiplication of an array by a matrix, whose product in general is an array. Since an array has three indices, it can be multiplied from three sides; we shall write

$$UX, \quad \overset{V}{X}, \quad XW$$

to indicate the three kinds of products; these indicate

$$\sum_i u_{vi} x_{ijk}, \quad \sum_i v_{ij} x_{ijk}, \quad \sum_k w_{k'k} x_{ijk}.$$

(It would be more in accord with matrix notation to write XW' for last product, but we shall not bother.) It is easy to see that this multiplication is associative and that $X = [A, B, C]$ implies

$$\overset{V}{UXW} = [UA, VB, WC].$$

From this it follows that

$$\text{rank}(UX) \leq \text{rank}(X).$$

If $\dim_1(UX) = \dim_1(X)$, there must be some matrix \tilde{U} such that $\tilde{U}UX = X$; therefore

$$\text{if } \dim_1(UX) = \dim_1(X), \text{ then } \text{rank}(UX) = \text{rank}(X).$$

Two useful auxiliary facts (*not* about array rank) hold when $X = [A, B, C]$:

$$\{u | uX = 0\} \supset \{u | uA = 0\}, \quad \text{where } u \text{ is a vector;}$$

$$\dim_1(X) \leq \text{rank}(A).$$

Let $w(y)$ = the weight of y = the number of non-zero elements in y . Then another useful fact is this:

$$\text{rank}(uX) \leq w(uA).$$

We shall call X 1-*nondegenerate* if $\dim_1(X)$ = the number of slabs X_i , and similarly for 2-nondegenerate and 3-nondegenerate.

To connect some results from complexity theory with our results, suppose that $\omega = (\omega_1, \dots, \omega_k)$ is a row vector of indeterminates (say commuting, though it makes no difference here). Then it is easy to interpret the row rank and column rank of the polynomial matrix $X\omega$ in terms of our concepts:

$$\text{row rank}(X\omega) = \dim_1(X), \quad \text{column rank}(X\omega) = \dim_2(X).$$

Using this, we can translate the row theorem of Fiduccia and the column theorem of Winograd into our terminology:

$$\text{length}(X) \geq \dim_1(X) \quad \text{and} \quad \text{length}(X) \geq \dim_2(X).$$

As stated in [9], these hold even when u_i and v_i are not indeterminates.

Now we present a generalization of Theorem 10 from Brockett and Dobkin [4], on whose paper this section of our paper draws heavily.

THEOREM 1. *If X is an array and U, V, W are matrices, then*

$$\text{rank}(X) \geq \dim_1(UX) + \text{rank}\left(\overset{V}{X}W\right) - \dim_1\left(\overset{V}{U}XW\right).$$

This generalizes a classical inequality in Frobenius [10] for *matrices* (see also Mirsky [15]), namely

$$\text{rank}(X) \geq \text{rank}(UX) + \text{rank}(XW) - \text{rank}(UXW).$$

If we set V to be the identity matrix so that it disappears from the array result, and set $J=1$ so X degenerates to an I -by- K matrix, then the analogy is essentially perfect. To specialize Theorem 1 to Brockett and Dobkin's theorem, let V be an identity matrix, let U consist of the last several rows of an identity matrix, let W consist of the first several rows of an identity matrix, assume $UXW=0$, and assume X is 1-nondegenerate, which implies that $\dim_1(UX)=\text{rank}(U)$. Note that UX then consists of the last several 1-slabs of X , XW consists of the first several 3-slabs of X , and UXW consists of the intersection of UX and XW , which is the lower left hand corner of X when we view its I -by- K face.

Brockett and Dobkin put their theorem to good use, as we describe below, which indicates *a fortiori* the value of Theorem 1. Simply to illustrate how it may be used, consider the array X in Fig. 1(b) as an example, and let W =identity matrix,

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then W may be ignored wherever it occurs, UX is the $\overset{V}{2}$ -by-4-by-4 array consisting of the upper two rows of the figure, and $\overset{V}{X}$ is the 4-by-2-by-4 array made up of the right-hand two columns of each slab in the figure. Also $\overset{V}{UX}$ is 2-by-2-by-4 and consists of the upper right-hand corner of each of the slabs in the figure; it is seen by inspection to be 0. Therefore by Theorem 1

$$\text{rank}(X) \geq \text{rank}(UXW) + \dim_2\left(\overset{V}{X}\right) - \dim_2\left(\overset{V}{UX}W\right) = 4 + 2 - 0 = 6,$$

since the elementary bounds show that

$$4 = \dim_3(UX) \leq \text{rank}(UX) \leq 4,$$

$$\dim_2\left(\overset{V}{X}\right) = 2.$$

Brockett and Dobdin consider the array which generalizes Fig. 1(b) to the case of multiplying an I -by- J matrix times a J -by- K matrix. This array X is IJ by JK by IK . It is convenient for the moment to denote each of the three subscripts by a pair of letters, so that the elements of X are indicated by $x_{(i_1, j_1), (j_2, k_1), (i_2, k_2)}$. It turns out that to describe multiplication of any other two matrices whose three sizes are I , J , and K , we get essentially the same array: that is, the second array may be obtained from the first by a combination of transposition and a permutation of slabs. For this reason, we may assume without real loss of generality that $I \geq J \geq K$. Elementary bounds yield that

$$\dim_1(X) \leq \text{rank}(X)$$

$$\leq (JK)(IK) - \text{card}\left\{\left((j_2, k_1), (i_2, k_2)\right) \mid x_{(i_1, j_1), (j_2, k_1), (i_2, k_2)} = 0 \text{ for all } (i_1, j_1)\right\},$$

so

$$IJ \leq \text{rank}(X) \leq IJK.$$

By using their result which corresponds to our Theorem 1 and induction, Brockett and Dobkin show that $\text{rank}(X) \geq I(J + K - 1)$. For the case $I = J = K$, this yields $\text{rank}(X) \geq 2I^2 - I$. For the case $I = J$ and $K \approx \log_2 I$ this yields

$$\text{rank}(X) \geq I^2 + I(\log_2 I - 1)$$

while in another paper [5] they show constructively that in this case

$$\text{rank}(X) \leq I^2 + O(I^2) \quad \text{for large } I.$$

Proof of Theorem 1. Let $R = \text{rank}(X)$ and let $X = [A, B, C]$ be a minimal representation, so (A, B, C) has rank R . Let I be the R -by- R identity matrix, and note that the R 1-slabs of the R -by- J -by- K array $[I, B, C]$ consist of R

dyads (matrices of rank 1) which are the outer products of the columns of B and the columns of C . Each of the I slabs of $[A, B, C]$ is a linear combination of these dyads, and the entries of A provide the coefficients, as we see from the identity

$$[A, B, C] = A[I, B, C].$$

Minimality of $[A, B, C]$ implies that $[I, B, C]$ is 1-nondegenerate. For suppose the contrary. Then there would be a linear relationship between the slabs of $[I, B, C]$, which would permit at least one of these slabs to be expressed as a linear combination of the others. Using this, the I slabs of $X = [A, B, C]$ could be expressed as linear combinations of $R - 1$ dyads, which cannot happen, since $\text{rank}(X) = R$.

Now define R_1 by

$$R_1 =_{\text{def}} \dim_1(UX) = \dim_1(UA[I, B, C]) = \text{rank}(UA),$$

where the last equality follows from the 1-nondegeneracy just proved. Pick R_1 independent rows of UA , and call the matrix they form U_1 . Then there is a "selection matrix" P_1 (that is, a matrix formed by a set of distinct rows of an identity matrix) and another matrix Q_1 such that

$$U_1 = P_1(UA), \quad UA = Q_1 U_1.$$

Now

$$\begin{aligned} R_2 &=_{\text{def}} \dim_1 \left(U \overset{V}{X} W \right) = \dim_1(UA[I, VB, WC]) \\ &= \dim_1(U_1[I, VB, WC]). \end{aligned}$$

Obviously

$$R_2 \leq R_1.$$

Thus $U_1[I, VB, WC]$ has R_1 slabs, which span a space of dimensionality R_2 . Thus there are $R_1 - R_2$ independent linear combinations of the slabs which are 0, that is, there is an $(R_1 - R_2)$ -by- R_1 matrix T of full row rank for which

$$TU_1[I, VB, WC] = 0.$$

Therefore

$$\dim_1[I, VB, WC] \leq R - \text{rank}(TU_1) = R - (R_1 - R_2),$$

so we can select some set of $R - R_1 + R_2$ slabs of $[I, VB, WC]$ which span the space they all span. Of course, the slabs are all dyads. Then

$$\overset{V}{X} W = A[I, VB, WC]$$

can be represented using just these dyads, so

$$\text{rank}\left(\overset{V}{X} W\right) \leq R - R_1 + R_2,$$

which is what we want to prove. ■

The next theorem is useful primarily through its corollaries.

THEOREM 2. *If X is a three-way array and $X = [A, B, C]$ is any representation, then*

$$R \geq \min_{T \text{ in } \mathfrak{T}} \text{rank}(TX) + \max_{S \text{ in } \mathfrak{S}} (\text{number of zero columns in } SA),$$

where \mathfrak{T} is any set of matrices and $\mathfrak{S} \subset \mathfrak{T}$.

COROLLARY 1. *Suppose X is 1-nondegenerate, and let $\mathfrak{T} = \{u|u \neq 0\}$, where u is a vector. Then*

$$\text{rank}(X) \geq \min_{u \text{ in } \mathfrak{T}} \text{rank}(uX) + \dim_1(X) - 1.$$

COROLLARY 1'. *Suppose X is 1-nondegenerate and $z \neq 0$, and let $\mathfrak{T} = \{u|u \cdot z \neq 0\}$. Then the same result holds.*

Corollary 1' reduces the size of the set over which the minimum must be taken in Corollary 1; probably still smaller sets can be used. Reasoning like that of Corollaries 1 and 1' may be found in [8] and [11], as well as [4].

Corollary 1 shows an interesting mathematical connection quite apart from anything else mentioned in this paper. The central concept in Berghman's work [2] is what he calls the "rank" of a linear space of matrices

(though he uses tensor rather than matrix terminology). He defines this to be the minimum rank of any nonzero matrix in the space. If we generalize the first term on the right-hand side of Corollary 1 slightly so we can drop the nondegeneracy assumption, we have $\min \text{rank}\{uX : \text{all } u \text{ with } uX \neq 0\}$, which is just Bergman's concept applied to the space generated by the 1-slabs of X .

COROLLARY 2. *Suppose X is 1-nondegenerate and \mathfrak{T} consists of all M -by- I matrices T with full row rank. Then*

$$\text{rank}(X) \geq \min_{T \text{ in } \mathfrak{T}} \text{rank}(TX) + \dim_1(X) - M.$$

COROLLARY 2*. *Suppose X is 1-nondegenerate and \mathfrak{T} consists of all M -by- I row-reduced echelon matrices. Then the same result holds.*

Corollary 2* reduces the size of the set over which the minimum must be taken; it is of interest in helping show the relationship between our results and those of others. Theorem 5 of Fiduccia and Zalcstein [9], which generalizes the row-column theorem of Fiduccia (see, e.g., Aho, Hopcroft, and Ullman [1, Chapter 12]), can be translated into our terminology as follows:

$$\text{length}(X) \geq \min_{T \text{ in } \mathfrak{T}} \dim_2(TX) + \dim_1(X) - M,$$

where \mathfrak{T} has the same meaning as in Corollary 2* and the variables in the bilinear form, u_i and v_j , need not be indeterminates (a point which Fiduccia and Zalcstein stress). Notice the similarity between this result and Corollary 2*. It may well be possible to strengthen this result by substituting $\text{length}(TX)$ for $\dim_2(TX)$.

Lemma 3.1 of van Leeuwen and van Emde Boas, which the authors refer to as the "crucial result" for their main theorem, can be stated as follows after many steps of translation and after interchanging the role of their x -variables and y -variables:

If Q is any permutation matrix of size I , and $0 \leq M \leq I$, and if $\dim_1((I_{I-M}, 0)QX) \geq I - M$, then

$$\text{length}(X) \geq \min_U \dim_2((U, I_M)QX) + I - M,$$

where U runs over all M -by- $(I-M)$ matrices, I is an identity matrix of indicated size, and the u_i and v_j are commuting indeterminates.

It is not hard to see that their method of proof yields the following stronger result, given the same hypotheses:

$$\text{length}(X) \geq \min_U \text{length}((U, I_M)QX) + I - M.$$

I believe the same method of proof can be slightly modified to yield the following result: if

$$0 \leq M \leq I,$$

then

$$\text{length}(X) \geq \min_{T \text{ in } \mathfrak{T}} \text{length}(TX) + I - M,$$

where \mathfrak{T} has the same meaning as in Corollary 2. This differs from Corollary 2 only in the substitution of "length" for "rank" and lack of the nondegeneracy assumption. [Because Corollary 2 has this assumption, we can change $\dim_1(X)$ to I there.]

COROLLARY 3. *If X is 1-nondegenerate and U is a given matrix, then*

$$\text{rank}(X) \geq \min_{\text{all } T} \text{rank}(X - TUX) + \text{rank}(U).$$

To specialize Corollary 3 to Theorem 9 of Brockett and Dobkin [4], set U equal to a matrix consisting of the last several rows of an identity matrix, and note that minimizing over T may be assumed to cancel the last several slabs of X in $X - TUX$. Obtaining their result as a specialization of ours appears to be simpler than their proof.

To illustrate the value of these corollaries, we use Corollary 1' to simplify the proofs of two already known results. The simpler one is Lemma 2 (p. 14) of Howell and Lafon [11]. This lemma states in effect that the rank of the following 3-by-3-by-4 array X is ≥ 6 :

-1								-1		1	
	-1				1				-1		
		-1		-1		1					

We use the version of Corollary 1' which yields

$$\text{rank}(X) \geq \min \text{rank}(Xu) + \dim_3(X) - 1.$$

$$Xu = \begin{bmatrix} -u_1 & u_4 & -u_3 \\ -u_4 & -u_1 & u_2 \\ u_3 & -u_2 & -u_1 \end{bmatrix},$$

Now we use Corollary 1' to give a brief proof of the chief result of de Groote [8]. This states that to compute both uv and vu where u and v are quaternions requires at least 10 real multiplications. (It was already known that 10 multiplications suffice.) If we express quaternions as 4-vectors in the usual way, then the array which describes the product $w = uv$ is this:

$$w$$

	v				v				v				v			
	1				1				1				1			1
	-1				1				-1				1			
u		-1				1	1			-1						
			-1			-1		1			1					

The first and fifth (u, v) -slices of this array are the same matrix, namely, $\text{diag}(1, -1, -1, -1)$, so the rank is not changed when we delete the fifth slice, and call the resulting 4-by-4-by-7 array X . Now we take $z = (1, 0, 0, 0, 0, 0, 0)$ and apply Corollary 1' to obtain

$$\text{rank}(X) \geq \min \text{rank}(Xt) + \dim_3(X) - 1,$$

where t is in $\mathfrak{T} = \{t | t_1 \neq 0\}$. It is not hard to see that $\dim_3(X) = 7$. To evaluate the first term we form Xt :

$$\begin{bmatrix} t_1 & t_2 + t_5 & t_3 + t_6 & t_4 + t_7 \\ t_2 + t_5 & -t_1 & t_4 - t_7 & -t_3 + t_6 \\ t_3 + t_6 & -t_4 + t_7 & -t_1 & t_2 - t_5 \\ t_4 + t_7 & t_3 - t_6 & -t_2 + t_5 & -t_1 \end{bmatrix}.$$

Now multiply the first row by -1 , which does not change the rank, and note that the symmetric part of the result is $t_1 \text{diag}(-1, -1, -1, -1)$, which is a definite matrix. As noted above, this proves that Xt has full rank, so $\min \text{rank}(Xt) = 4$. Then $\text{rank}(X) \geq 4 + 7 - 1 = 10$.

Proof of Theorem 2. We have for any S in \mathfrak{S} ,

$$\begin{aligned} \min_{T \text{ in } \mathfrak{T}} \text{rank}(TX) &\leq \text{rank}(SX) = \text{rank}([SA, B, C]) \\ &\leq R - (\text{number of zero columns in } SA). \end{aligned}$$

If we now take the minimum over all S in \mathfrak{S} , we get the Theorem. ■

Proof of Corollary 1. Pick some representation $[A, B, C]$ of X . Let $\mathfrak{S} = \{u | uA \neq 0\}$. Then

$$\mathfrak{T} = \{u | uX \neq 0\} \subset \mathfrak{S} \subset \mathfrak{T}$$

where the equality comes from the 1-nondegeneracy of X , the first inclusion comes from the elementary facts, and the last inclusion is obvious. Thus $\mathfrak{S} = \mathfrak{T}$. (Also, for use in proving Corollary 1', $\text{rank}(A) = \text{number of rows of } A$.) Using "max" to indicate the maximum over all u in \mathfrak{S} , we can evaluate the last term in Theorem 2:

$$\begin{aligned} &\max(\text{number of zero columns in } uA) \\ &= \max(\text{number of columns of } A \text{ which are orthogonal to } u) \\ &\geq \text{rank}(A) - 1 \geq \dim_1(X) - 1. \end{aligned}$$

The first inequality follows because we can pick some $\text{rank}(A)$ independent columns of A , and select u orthogonal to $\text{rank}(A) - 1$ of them. Corollary 1 is now immediate from Theorem 2. ■

Proof of Corollary 1'. Let $\mathcal{S} = \{u | uA \neq 0 \text{ and } u \cdot z \neq 0\}$. Then

$$\mathcal{T} = \{u | uX \neq 0 \text{ and } u \cdot z \neq 0\} \subset \mathcal{S} \subset \mathcal{T}.$$

Here the equality follows because u in \mathcal{T} implies $u \neq 0$, and X 1-nondegenerate then implies $uX \neq 0$. The first inclusion follows from the elementary facts, and the other inclusion is trivial. Thus $\mathcal{S} = \mathcal{T}$. We can then proceed as in the proof of Corollary 1, but the first inequality in the chain needs further explanation.

From the proof of Corollary 1 we see that the columns of A span the space of all possible column vectors. Now we pick some $\text{rank}(A)$ independent columns of A , and note that they form a basis. We form the dual basis, and note that at least one element of it is not orthogonal to z . We choose u to be this element, and the proof is complete. ■

Proof of Corollaries 2 and 2.* Let $\mathcal{S} = \mathcal{T}$. Using "max" to indicate the maximum over all S in \mathcal{S} , we can evaluate the last term in Theorem 2:

$$\begin{aligned} & \max(\text{number of zero columns in } SA) \\ &= \max(\text{number of columns of } A \text{ which are orthogonal to all } M \text{ rows of } S) \\ &\geq \text{rank}(A) - M \geq \dim_1(X) - M. \end{aligned}$$

For Corollary 2, the first inequality follows because we can select any $\text{rank}(A) - M$ independent columns of A , and choose the rows of S to be any M independent vectors which are orthogonal to the selected columns of A .

For Corollary 2*, we obtain the matrix S in the same manner and then perform row operations to reduce it to row-reduced echelon form. ■

Proof of Corollary 3. Let I be the I -by- I identity matrix, and let $M = \text{rank}(U)$. Let $\mathcal{T} = \mathcal{S} = \{I - TU | \text{all } T\}$. Then we only need to show that $(I - TU)A = A - TUA$ has at least M zero columns for suitable T . Since X is 1-nondegenerate, A has full row rank, so $\text{rank}(UA) = M$. Therefore UA contains a set of M columns \bar{U} of rank M , and these contain a square nonsingular submatrix \tilde{U} of order M . Let \bar{A} be the submatrix of A containing the M columns which correspond to \bar{U} . Then there is a matrix T_1 such that $T_1 \tilde{U} = \bar{A}$. Now let P consist of M rows from I , so selected that $P(UA)$ contains \tilde{U} . Then $T_1 P(UA)$ agrees with A in the columns of \bar{A} , so $A - T_1 P(UA)$ is zero in M columns. ■

III. MORE LOWER BOUNDS ON RANK: THEOREMS 3a-d

Given one known triple product representation of an array X , it is possible to give a lower bound on $\text{rank}(X)$ in terms of that representation. This section is devoted to proving several results of this sort. These results are primarily of interest in the data-analysis context, where we start with an array X and fit an *approximate* triple product decomposition. If we let \hat{X} be the triple product of the three matrices involved, then in most situations we would expect the conditions of Theorem 3a to hold with probability 1 for \hat{X} . Typically we have $I_0 + J_0 + K_0 \geq 2R + 1$, so that Theorem 3a yields that the rank of the representation is $\text{rank}(\hat{X})$. In many cases where Theorem 3a does not apply in this way, the other theorems can be applied instead, to yield the same result. This reassures us that \hat{X} cannot be expressed in terms of a lower rank representation. If \hat{X} is in fact the least-squares approximation to X of its rank (which usually appears to be the case in practice), this reassures us that an approximation to X which fits as well as \hat{X} is unlikely to be available with lower rank. These simple reassurances are of course important to the practical use of the statistical method.

The four results involved form a series, in which each succeeding result is more general than the preceding one, but harder to use. All the results follow from a single uniform line of proof in which we demonstrate that

$$\text{Theorem 3a} \Leftarrow \text{Theorem 3b} \Leftarrow \text{Theorem 3c} \Leftarrow \text{Theorem 3d}$$

and finally prove Theorem 3d.

We make the following assumptions throughout this section, and for all four theorems in it. First, we assume that $X = [A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$, where the square brackets indicate the triple product of matrices, and where the two representations have rank R and \bar{R} respectively. We let $I_0 = \text{rank}(A)$, $J_0 = \text{rank}(B)$, and $K_0 = \text{rank}(C)$. We assume that X is not identically zero, and that none of the columns of A, B, C are zero. We assume that the latter representation has minimum possible rank, so that $\bar{R} = \text{rank}(X)$ and $R \geq \bar{R}$. P always indicates a permutation matrix and Λ a diagonal matrix.

As the four theorems are stated, the second and third factors, B and C , play symmetric roles, but A enters the theorem in a special way. In view of the general symmetry among the three factors which we described earlier in the paper, each of these theorems has two other versions, in which the second or third factor plays the special role. However, we leave formulation of these alternative versions to the reader.

Note the trivial fact that $I_0 \leq R$, $J_0 \leq R$, $K_0 \leq R$.

THEOREM 3a. *Suppose every J_0 columns of B are linearly independent and every K_0 columns of C are linearly independent. If $J_0 + K_0 \geq R + 1$, then*

$$\bar{R} \geq I_0 + \min(R - I_0, J_0 - 1, K_0 - 1, J_0 + K_0 - R - 1) \geq I_0.$$

If in addition, $I_0 + J_0 + K_0 \geq 2R + 1$, then $\bar{R} = R$, and $\bar{A} = AP\Lambda$ for some P and Λ .

Since B has rank J_0 , it contains some set of J_0 columns which are linearly independent. To assume that *every* set of J_0 columns is independent, however, says a good deal more. In the next theorem we weaken this condition.

THEOREM 3b. *Suppose there are numbers $J_1 \leq J_0$ and $K_1 \leq K_0$ such that every J_1 columns of B are linearly independent and every K_1 columns of C are linearly independent. If*

$$J_0 + K_0 - \min(J_0 - J_1, K_0 - K_1) \geq R + 1,$$

$$\min(J_1, K_1) \geq 1,$$

then

$$\bar{R} \geq I_0 + \min(R - I_0, J_1 - 1, K_1 - 1,$$

$$J_0 + K_0 - R - 1 - \min(J_0 - J_1, K_0 - K_1)) \geq I_0.$$

If in addition

$$I_0 + J_0 + K_0 - \min(J_0 - J_1, K_0 - K_1) \geq 2R + 1,$$

$$\min(J_1, K_1) + I_0 \geq R + 1,$$

then $\bar{R} = R$, and $\bar{A} = AP\Lambda$ for some P and Λ .

The next theorem weakens the condition still further. To state it requires the use of some notation to indicate the rank of any set of columns of B and

of C . Let \tilde{B} indicate a set of columns of B and \tilde{C} indicate a set of columns of C . Let

$$r_B(\delta) = \min_{\text{card}(\tilde{B}) = \delta} \text{rank}(\tilde{B}), \quad \text{similarly for } r_C,$$

and

$$h(\delta) =_{\text{def}} r_B(\delta) + r_C(\delta) - \delta.$$

THEOREM 3c. *If for some $h_0 \geq 1$ we have*

$$h(\delta) \geq \min(\delta, h_0) \quad \text{for all } \delta,$$

then

$$\bar{R} \geq \min(R, I_0 + h_0 - 1) \geq I_0.$$

If in addition, $h_0 \geq R - I_0 + 1$, then $\bar{R} = R$, and $\bar{A} = AP\Lambda$ for some P and Λ .

To weaken the condition still further, we introduce the function

$$H(\delta) =_{\text{def}} \min_{\text{card}(\tilde{B}) = \delta} [\text{rank}(\tilde{B}) + \text{rank}(\tilde{C}) - \delta],$$

where \tilde{B} is any set of columns of B , and \tilde{C} is the *corresponding* set of columns of C .

THEOREM 3d. *If for some $H_0 \geq 1$ we have*

$$H(\delta) \geq \min(\delta, H_0) \quad \text{for all } \delta,$$

then

$$\bar{R} \geq \min(R, I_0 + H_0 - 1) \geq I_0.$$

If in addition, $H_0 \geq R - I_0 + 1$, then $\bar{R} = R$, and $\bar{A} = AP\Lambda$ for some P and Λ .

Consider the array shown in Fig. 1(b), and the representation of it shown in Fig. 2(b). Solely to illustrate the theorems above, and to make their meaning clear, we shall apply them to this situation. (The results we get are

weaker than the already known fact that the array has rank 7. This follows because it has been proved several times that the minimum number of scalar multiplications needed to form the product of two matrices is 7.) We have $R=7$, and it is easy to see that $I_0=J_0=K_0=4$. Theorem 3a does not apply because, for example, columns 1, 2, 4 of B are dependent. It is easy to check that $I_1=J_1=K_1=2$ are the largest possible values but Theorem 3(b) does not apply because the main numerical condition does not hold:

$$J_0 + K_0 - \min(J_0 - J_1, K_0 - K_1) = 4 + 4 - 2 \not\geq 7 + 1.$$

However, Theorem 3c does apply. To see this requires some analysis of the combinatorial geometries (or matroids) consisting of the columns of each of A , B , C . Fortunately, these geometries are all isomorphic, since

$$A = D_B B D_B^* P_B = D_C C D_C^* P_C,$$

where each D is a suitable diagonal matrix with ± 1 on the diagonals and each P is a suitable permutation matrix. Thus much of our analysis need be done for A only.

Let us denote the seven columns of A by the following shorthand,

$$1, \quad 4, \quad 14, \quad 12, \quad 24, \quad 13, \quad 34,$$

where ij indicates a column which is non-zero in positions i and j , and so forth. It is easy to verify that the circuits of A (that is, the minimal dependent sets) consist precisely of the following six sets, which have a very simple structure:

$$\begin{array}{llll} \{14, & 1, 4\}, & & \\ \{14, & & 12, 24\}, & \\ \{14, & & & 13, 34\}, \\ & \{1, 4 & 12, 24\}, & \\ & \{1, 4 & & 13, 34\}, \\ & & \{12, 24, & 13, 34\}. \end{array}$$

Each circuit is the union of precisely two of the following sets:

$$\{14\}, \quad \{1, 4\}, \quad \{12, 24\}, \quad \{13, 34\}.$$

With the aid of these facts, it is easy to verify that the rank function can be described as in Table 1. Let

$$h(\delta) \equiv r_B(\delta) + r_C(\delta) - \delta,$$

TABLE 1

δ	$r_A(\delta)$	Rank of Set of Size δ	
0	0	0	
1	1	1	
2	2	2	
3	2	3	unless set is circuit
		2	otherwise
4	3	4	unless set contains a circuit
		3	otherwise
5	3	3	if it is the complement of
		4	{1, 4}, of {12, 24}, or of {13, 34}
6	4	4	otherwise
7	4	4	

and in this situation

$$h(\delta) = 2r_A(\delta) - \delta.$$

Then we get the values of h to be successively 0, 1, 2, 1, 2, 1, 2, 1. From this we find that the best $h_0 = 1$, so Theorem 3c applies. It yields

$$\overline{R} \geq 4 + \min(1, 0) = 4,$$

which is no better than the elementary bounds.

Theorem 3d also applies, but unfortunately does not give a better result. It is not hard to verify, after expressing the permutation connecting the columns of B with the columns of C in a simple diagram, that, the values of

$$H(\delta) = \min_{\tilde{B}} [\text{rank}(\tilde{B}) + \text{rank}(\tilde{C}) - \delta], \quad \text{where } \text{card}(\tilde{B}) = \delta,$$

are successively 0, 1, 2, 2, 2, 2, 2, 1: this calculation is aided by the obvious facts that $H(\delta) \geq h(\delta)$ and $H(R) = h(R)$. If $H(7)$ were 2 instead of 1, we would get a lower bound of 5 instead of 4 for \overline{R} .

Proof that Theorem 3b \Rightarrow Theorem 3a. Note that since X is assumed not identically 0, $\min(I_0, J_0, K_0) \geq 1$. To obtain Theorem 3a from Theorem 3b, we merely set $J_1 = J_0$ and $K_1 = K_0$. ■

Proof that Theorem 3c \Rightarrow Theorem 3b. We assume the hypotheses of Theorem 3b, and prove that the hypotheses of Theorem 3c follow from

them. This is sufficient, for the conclusion of Theorem 3c easily yields that of Theorem 3b. For this purpose we introduce functions g_B , g_C , and h^* analogous to r_B , r_C , and h of Theorem 3c. It will turn out that each new function is less than or equal to its corresponding function. Let $J_3 = J_1 - J_0 + R$, note that $J_1 \leq J_3 \leq R$, and define

$$g_B(\delta) = \begin{cases} \delta & \text{if } 0 \leq \delta \leq J_1, \\ J_1 & \text{if } J_1 \leq \delta \leq J_3, \\ \delta + J_0 - R & \text{if } J_3 \leq \delta \leq R. \end{cases}$$

Note that g_B is continuous and piecewise linear with slope $+1, 0, +1$ successively in the three pieces. We claim $r_B(\delta) \geq g_B(\delta)$ for all δ . For δ in the first two pieces, this is an elementary consequence of the assumption that every J_1 columns of B are linearly independent. For δ in the third piece, it follows because B has rank J_0 and because removing one column never reduces the rank by more than one. Making similar definitions and observations for g_C , we now have

$$h(\delta) \geq h^*(\delta) \equiv_{\text{def}} g_B(\delta) + g_C(\delta) - \delta.$$

Now h^* is continuous and piecewise linear with breakpoints at J_1, J_3, K_1, K_3 . The slope of the first piece is $+1$, and it goes from 0 to $\min(J_1, K_1)$. Thus $h^*(\delta) = \delta$ for $\delta \leq \min(J_1, K_1)$, and hence the assumption in Theorem 3c will hold for

$$h_0 = \min_{\delta \geq \min(J_1, K_1)} h^*(\delta).$$

To prove that $h_0 \geq 1$, we consider two cases. If either $J_3 \leq K_1$ or $K_3 \leq J_1$, the slopes of the five pieces of h^* are easily seen to be $+1, 0, +1, 0, +1$ successively. In this case h^* is weakly increasing, so its minimum occurs at the left endpoint, so $h_0 \geq \min(J_1, K_1)$, and by assumption this is ≥ 1 .

If $J_3 > K_1$ and $K_3 > J_1$, then the four breakpoints of h^* are, in order of increasing size,

$$\min(J_1, K_1), \quad \max(J_1, K_1), \quad \min(J_3, K_3), \quad \max(J_3, K_3),$$

and the slopes of the five pieces are $+1, 0, -1, 0, +1$ successively. From this it is clear that the minimum value of h^* for $\delta \geq \min(J_1, K_1)$ occurs during the latter of the pieces with slope 0 , and in particular at $\max(J_3, K_3)$. Thus

$h_0 \geq h^*(\max(J_3, K_3))$. But for $\delta \geq \max(J_3, K_3)$ we have

$$\begin{aligned} h^*(\delta) &= [\delta + J_0 - R] + [\delta + K_0 - R] - \delta \\ &= J_0 + K_0 - 2R + \delta, \\ h_0 &\geq J_0 + K_0 - 2R + \max(J_1 - J_0 + R, K_1 - K_0 + R) \\ &= J_0 + K_0 - R - \min(J_0 - J_1, K_0 - K_1). \end{aligned}$$

By the assumption in Theorem 3b this is ≥ 1 , so in this case also $h_0 \geq 1$, and the condition of Theorem 3c is thereby satisfied. Now we combine the conclusion of Theorem 3c with that we have just proved about h_0 , namely

$$h_0 \geq \min(J_1, K_1, J_0 + K_0 - R - \min(J_0 - J_1, K_0 - K_1)),$$

and we get the conclusion of Theorem 3b. ■

Proof that Theorem 3d \Rightarrow Theorem 3c. Note that

$$\begin{aligned} H(\delta) &\geq \min \text{rank}(\tilde{B}) + \min \text{rank}(\tilde{C}) - \delta \\ &= r_B(\tilde{B}) + r_C(\tilde{C}) - \delta = h(\delta). \end{aligned}$$

Therefore $H_0 \geq h_0$. From the assumption of Theorem 3c we have $h_0 \geq 1$, hence $H_0 \geq 1$, and the conclusion of Theorem 3d may be used, from which the conclusion of Theorem 3c immediately follows. ■

Proof of Theorem 3d. Our basic assumption, of course, is $[A, B, C] = X = [\bar{A}, \bar{B}, \bar{C}]$. Denote the matrix comprising the i th slab of X by X_i , and let $a_{i\rightarrow}$ indicate the i th row of A . Then it is easy to see that

$$B \text{diag}(a_{i\rightarrow})C' = X_i = \bar{B} \text{diag}(\bar{a}_{i\rightarrow})\bar{C}',$$

where C' indicates the transpose of C , and $\text{diag}(y)$ for any vector y indicates a diagonal matrix with the values of y running down the diagonal. If x is any row vector with I entries, then by taking linear combinations of slabs we find that

$$B \text{diag}(xA)C' = \bar{B} \text{diag}(x\bar{A})\bar{C}'.$$

This is the form in which we use the basic assumption.

Our first step is to prove a lemma.

RANK LEMMA. Suppose D is any diagonal R -by- R matrix of rank δ . Then

$$\text{rank}(BDC') \begin{cases} = \delta & \text{for } \delta \leq H_0, \\ \geq H_0 & \text{for } \delta \geq H_0. \end{cases}$$

Proof. Let D_1 be the diagonal matrix with 1's where D has nonzero elements, and 0's elsewhere. Then $(BD)(CD_1)' = BDC'$. BD and CD_1 have columns of 0's corresponding to the zero elements on the diagonal of D . Form \tilde{B} and \tilde{C} by dropping these zero columns. Then \tilde{B} and \tilde{C} have δ columns each, and $\tilde{B}\tilde{C}' = BDC'$. Now we recall the well-known fact that if Y and Z are matrices with δ columns each, then

$$\min[\text{rank}(Y), \text{rank}(Z)] \geq \text{rank}(YZ') \geq \text{rank}(Y) + \text{rank}(Z) - \delta.$$

Applying this to $\tilde{B}\tilde{C}'$, we get

$$\begin{aligned} \delta &\geq \min(\text{rank}(\tilde{B}), \text{rank}(\tilde{C})) \geq \text{rank}(BDC') \\ &\geq \text{rank}(\tilde{B}) + \text{rank}(\tilde{C}) - \delta \geq H(\delta). \end{aligned}$$

If $\delta \leq H_0$, then $H(\delta) \geq \delta$ by the hypothesis of Theorem 3d, and this gives the first case in the lemma. If $\delta \geq H_0$, then $H(\delta) \geq H_0$ for the same reason, and this gives the second case. ■

Now let $\text{col}(A)$ indicate the column space of A and $\text{null}(A) = \{x | xA = 0\}$ be its orthogonal complement. Then we claim that

$$\text{null}(A) \supset \text{null}(\bar{A}),$$

$$\text{col}(A) \subset \text{col}(\bar{A}),$$

$$I_0 \leq \bar{I}_0 =_{\text{def}} \text{rank}(\bar{A}).$$

It suffices to show the first of these. If x is in $\text{null}(\bar{A})$, then $x\bar{A} = 0$, so

$$B \text{diag}(xA)C' = \bar{B} \text{diag}(x\bar{A})\bar{C}' = 0,$$

so $\text{rank } [B \text{diag}(xA)C'] = 0$. Since $H_0 \geq 1$, the Rank Lemma shows that $\text{rank diag}(xA) = 0$, so $xA = 0$, as desired.

Let $w(y)$ indicate the number of nonzero entries in y . If $w(x\bar{A}) \leq \bar{R} - \bar{I}_0 + 1$, then we claim that $w(xA) \leq w(x\bar{A})$. (This is the condition we need in order to be able to apply the Permutation Lemma.) First, we have

$$\begin{aligned} R - I_0 + 1 &\geq \bar{R} - \bar{I}_0 + 1 \geq w(x\bar{A}) = \text{rank diag}(x\bar{A}) \\ &\geq \text{rank} [\bar{B} \text{diag}(x\bar{A}) \bar{C}'] = \text{rank} [B \text{diag}(xA) C'] \\ &\begin{cases} = \text{rank diag}(xA) = w(xA) & \text{if } w(xA) \leq H_0, \\ \geq H_0 & \text{otherwise,} \end{cases} \end{aligned}$$

using the Rank Lemma and the fact that the rank of a matrix product is no greater than the rank of any factor. If $\bar{R} \geq I_0 + H_0 - 1$, then the conclusion of Theorem 3d already holds, so suppose to the contrary that $\bar{R} \leq I_0 + H_0 - 2$. This makes the second case above impossible, so the first case must hold, so that

$$w(xA) \leq w(x\bar{A}).$$

Now we invoke the Permutation Lemma, which is stated and proved in a separate section. All assumptions needed there have been assumed or proved. The lemma shows that $R = \bar{R}$ and that $\bar{A} = AP\Lambda$ as desired. ■

4. UNIQUENESS OF TRIPLE PRODUCT DECOMPOSITIONS: THEOREM 4

The bilinear statistical methods (factor analysis and principal components) involve expressing a *matrix* X as an (ordinary matrix) product AB' (approximately). Because $AB' = (AT)(BT'^{-1})'$ for any nonsingular T , any factorization gives rise to many others of the same rank, and leads to a selection problem. In factor analysis this is called the rotation problem, and it consumes considerable attention and effort. By comparison, individual differences scaling, through its use of canonical decomposition, relies on expressing an array X (approximately) as a triple product $[A, B, C]$ of matrices. As we show in this section, the alternative representations are much more limited: in most cases of interest in data analysis, the only alternative

triple-product representations (of the same rank) have the form

$$[AP\Lambda, BPM, CPN]$$

where P is a *permutation* matrix, and Λ , M , N are *diagonal* matrices with $AMN =$ the identity matrix. Thus the coordinate system has a special status, and free rotation is *not* possible: this is of great practical significance.

We prove a series of six uniqueness theorems. The conclusion is always the same, but the hypotheses become increasingly weak and increasingly difficult to apply. The four theorems labeled 4a, 4b, 4c, 4d correspond to the four correspondingly labeled theorems in the previous section. Theorem 4i states a special elementary result which holds only for the case $R=1$. In most typical statistical situations, the hypotheses of one of these theorems would apply, often those of Theorem 3a.

We make the following assumptions throughout this section and in all six theorems. The chief assumption is that $[A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$. We assume that each of the six matrices has R columns. We let $I_0 = \text{rank}(A)$, $J_0 = \text{rank}(B)$, $K_0 = \text{rank}(C)$. P is always a permutation matrix, and Λ , M (μ), and N (ν) are always diagonal matrices. We call (A, B, C) *equivalent* to $(\bar{A}, \bar{B}, \bar{C})$ if there exist P, Λ, M, N such that

$$\bar{A} = AP\Lambda, \quad \bar{B} = BPM, \quad \bar{C} = CPN, \quad AMN = \text{identity}.$$

This is in fact an equivalence relation among triples of matrices (that is, it is reflexive, symmetric, and transitive).

THEOREM 4i. *If $R=1$ and $I_0 + J_0 + K_0 \geq 3$, then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.*

Since $\text{rank}(A) = I_0$, some set of I_0 columns of A must be independent, but it is much stronger to assume that every set is independent.

THEOREM 4a. *Suppose every I_0 columns of A are independent, every J_0 columns of B are independent, and every K_0 columns of C are independent. Suppose $I_0 + J_0 + K_0 \geq 2R + 2$. Then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.*

In the following theorems the first factor A has a different role than the second and third factors, B and C . Of course each theorem has two other versions, in which the second or the third factor plays the special role, but we leave the formulation of these versions to the reader.

THEOREM 4b. *Suppose there are numbers $I_1 \leq I_0$, $J_1 \leq J_0$, $K_1 \leq K_0$ such that every I_1 columns of A are independent, every J_1 columns of B are*

independent, and every K_1 columns of C are independent. Suppose the following conditions are satisfied:

$$\min(I_1, K_1) + J_0 \geq R + 2,$$

$$\min(I_1, J_1) + K_0 \geq R + 2,$$

$$\left. \begin{array}{l} \min(I_0 - I_1, J_0 - J_1) \\ \min(I_0 - I_1, K_0 - K_1) \end{array} \right\} \leq I_0 + J_0 + K_0 - (2R + 2).$$

Then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.

Let \tilde{A} be a set of columns of A , and define

$$r_A(\delta) = \min_{\text{card}(\tilde{A}) = \delta} \text{rank}(\tilde{A}).$$

Define r_B and r_C similarly. Note that each of these functions satisfies $r(\delta) \leq r(\delta + 1) \leq r(\delta) + 1$. Define $h_{AB}(\delta) = r_A(\delta) + r_B(\delta) - \delta$, and define h_{AC} and h_{BC} similarly. Note that each function h satisfies $h(\delta) \leq \delta$ and $|h(\delta) - h(\delta + 1)| \leq 1$.

THEOREM 4C. Suppose I_1, J_1, K_1 have the same properties as in Theorem 4b. Also suppose the following conditions are satisfied:

$$I_1 \geq \max(R - J_0 + 2, R - K_0 + 2),$$

$$J_1 \geq R - K_0 + 2,$$

$$K_1 \geq R - J_0 + 2,$$

$$h_{AB}(\delta) \geq \min(\delta, R - K_0 + 2),$$

$$h_{AC}(\delta) \geq \min(\delta, R - J_0 + 2),$$

$$h_{BC}(\delta) \geq \min(\delta, 1).$$

Then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.

Let \tilde{B} be any set of columns of B , let \tilde{C} be the corresponding set of columns of C , and define

$$H_{BC}(\delta) = \min_{\text{card}(\tilde{B}) = \delta} [\text{rank}(\tilde{B}) + \text{rank}(\tilde{C}) - \delta].$$

Define H_{AB} and H_{AC} in a similar manner. Note that each of these three functions satisfies $H(\delta) \leq \delta$ and $|H(\delta) - H(\delta + 1)| \leq 1$.

THEOREM 4d. *Suppose I_1, J_1, K_1 have the same properties as in Theorem 4b. Also suppose the following conditions are satisfied:*

$$I_1 \geq \max(R - J_0 + 2, R - K_0 + 2),$$

$$J_1 \geq R - K_0 + 2,$$

$$K_1 \geq R - J_0 + 2,$$

$$H_{AB}(\delta) \geq R - K_0 + 2 \quad \text{if } \delta \geq R - K_0 + 2,$$

$$H_{AC}(\delta) \geq R - J_0 + 2 \quad \text{if } \delta \geq R - J_0 + 2,$$

$$H_{BC}(\delta) \geq 1 \quad \text{if } \delta \geq 1.$$

Then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.

THEOREM 4e. *Suppose the following conditions are satisfied:*

$$H_{AB}(\delta) \geq \min(\delta, R - K_0 + 2),$$

$$H_{AC}(\delta) \geq \min(\delta, R - J_0 + 2),$$

$$H_{BC}(\delta) \geq \min(\delta, 1).$$

Then (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are equivalent.

Proof of Theorem 4i. Since $R = 1$, the six matrices A, B, \dots each consist of a column matrix, and any permutation matrix is the 1-by-1 identity matrix. The other hypothesis yields that $I_0 = J_0 = K_0 = 1$, so none of the six matrices consists of a zero vector. It is then trivial to prove that \bar{A} is proportional to A , and similarly for B and C , and we have the desired result. ■

For the remaining theorems we first prove that

$$\begin{aligned} \text{Theorem 4a} &\Leftarrow \text{Theorem 4b} \Leftarrow \text{Theorem 4c} \\ &\Leftarrow \text{Theorem 4d} \Leftrightarrow \text{Theorem 4e.} \end{aligned}$$

Note that the final arrow is double-ended. Then we prove that Theorems 4d and 4e are true using the hypotheses of both, which is legitimate because of the equivalence.

Proof that Theorem 4b \Rightarrow Theorem 4a. Assuming the hypotheses of Theorem 4a, we may set the parameters I_1, J_1, K_1 of Theorem 4b equal to I_0, J_0, K_0 . By combining the inequality $I_0 + J_0 + K_0 \geq 2R + 2$ from Theorem 4a with the trivial inequalities $I_0 \leq R, J_0 \leq R, K_0 \leq R$ we obtain the three inequalities

$$I_0 + J_0 \geq R + 2, \quad I_0 + K_0 \geq R + 2, \quad J_0 + K_0 \geq R + 2,$$

and from these the first two inequalities in Theorem 4b follow immediately. The final pair of inequalities in Theorem 4b follow even more easily, so we may apply Theorem 4b and obtain the desired conclusion. ■

Proof that Theorem 4c \Rightarrow Theorem 4b. We note that the first three inequalities in Theorem 4c follow easily from those in Theorem 4b. Proving the remaining three inequalities takes more work. We note that $r_A(\delta) \leq \delta$ for all δ , and similarly for r_B and r_C , so that

$$h_{AB}(\delta) \leq \delta \quad \text{for all } \delta.$$

For $\delta \leq \min(I_1, J_1)$ we have $h_{AB}(\delta) = \delta$, and since $\min(I_1, J_1) \geq R - K_0 + 2$, we have proved part of the inequality on h_{AB} . To prove the rest, we let $I_2 = I_1 - I_0 + R$, and note that it is very easy to prove the following inequality:

$$r_A(\delta) \geq g_A(\delta) =_{\text{def}} \begin{cases} \delta & \text{if } 0 \leq \delta \leq I_1, \\ I_1 & \text{if } I_1 \leq \delta \leq I_2, \\ \delta + I_0 - R & \text{if } I_2 \leq \delta \leq R. \end{cases}$$

For $\delta \leq I_2$ this inequality follows directly from the assumed property of I_1 . For $\delta \geq I_2$, it follows because A has rank I_0 and because removing one column cannot reduce rank by more than 1. Similarly we define g_B and g_C , and get similar inequalities.

The function g_A is continuous and piecewise linear with two breakpoints, and has slopes $+1, 0, +1$ in the three pieces. Then

$$h_{AB}(\delta) \geq h_{AB}^*(\delta) =_{\text{def}} g_A(\delta) + g_B(\delta) - \delta,$$

and h_{AB}^* is continuous and piecewise linear with four breakpoints (namely, I_1, I_2, J_1, J_2). If these breakpoints occur in the order mentioned, or with both

J 's preceding both I 's, then the slopes of the five pieces are $+1, 0, +1, 0, +1$. If the breakpoints occur in any of the four other possible orders, the slopes are $+1, 0, -1, 0, +1$. Now we calculate the minimum value of $h_{AB}^*(\delta)$ for $\delta \geq R - K_0 + 2$ in both cases. In the first case it occurs (among other places) at $\min(I_1, J_1)$, and the value of h_{AB}^* there is $\min(I_1, J_1)$. In the second case it occurs (among other places) at $\max(I_2, J_2)$, and the value of h_{AB}^* there is

$$I_0 + J_0 - R - \min(I_0 - I_1, J_0 - J_1).$$

By the inequalities in Theorem 4b we find that the minimum values in both cases are $\geq R - K_0 + 2$. This proves the inequality involving r_A and r_B . The remaining two inequalities are proved in similar fashion, so Theorem 4c may be applied to reach the desired conclusion. ■

Proof that Theorem 4d \Rightarrow Theorem 4c. This is almost trivial. We see directly from the definitions that

$$H_{AB}(\delta) \geq r_A(\delta) + r_B(\delta) - \delta,$$

so the inequalities in Theorem 4c imply those in Theorem 4d, and the latter theorem may be applied to reach the desired conclusion. ■

Proof that Theorem 4e \Leftrightarrow Theorem 4d. It is enough to show that the inequalities of the two theorems are equivalent. For this purpose we need the elementary properties of the functions H which were mentioned at the time of definition. First we assume the inequalities in Theorem 4d. Together with the elementary properties, the last three of these inequalities show that

$$H_{AB}(\delta) = \delta \quad \text{for } 0 \leq \delta \leq R - K_0 + 2,$$

$$H_{AC}(\delta) = \delta \quad \text{for } 0 \leq \delta \leq R - J_0 + 2,$$

$$H_{BC}(\delta) = \delta \quad \text{for } 0 \leq \delta \leq 1.$$

Combining these with the same inequalities just used yields the inequalities of Theorem 4e.

Next we assume the inequalities of Theorem 4e. These immediately yield the last three inequalities of Theorem 4d. Together with the elementary properties of the functions H they also yield the equations in the preceding paragraph. Looking back at the definition of H_{AB} , and looking at the first of

these equations, we see that

$$\text{rank}(\tilde{B}) + \text{rank}(\tilde{C}) - \delta \geq H_{AB}(\delta) = \delta \quad \text{for } 0 \leq \delta \leq R - K_0 + 2,$$

whenever \tilde{B} is a set of δ columns of B , and \tilde{C} is the corresponding set of columns of C . Then using the trivial inequalities $\delta \geq \text{rank}(\tilde{B})$ and $\delta \geq \text{rank}(\tilde{C})$, we find that $\text{rank}(\tilde{B}) = \delta$ and $\text{rank}(\tilde{C}) = \delta$. Thus every set of columns in B and C up to size $R - K_0 + 2$ is independent, so

$$J_1 \geq R - K_0 + 2, \quad K_1 \geq R - K_0 + 2.$$

Using similar arguments for the other three equations, we obtain the first three inequalities of Theorem 4d. ■

Proof of Theorems 4d and 4e. We assume the inequalities of both theorems, and prove their common conclusion. From these inequalities and the trivial inequalities $R \geq I_0$, $R \geq J_0$, $R \geq K_0$ we derive $I_1 \geq 2$, $J_1 \geq 2$, $K_1 \geq 2$. We recall the chief assumption, stated earlier in this section but not explicitly used until now, that $[A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$. As in the proof of Theorem 3d in an earlier section, this yields the fact that

$$B \text{diag}(xA)C' = \bar{B} \text{diag}(x\bar{A})\bar{C}'$$

for any x .

Let $\text{col}(C)$ indicate the column space of C , and let $\text{null}(C) = \{x | xC = 0\}$ be its orthogonal complement. Let $\bar{K}_0 = \text{rank}(\bar{C})$. Then we claim

$$\text{null}(C) \supset \text{null}(\bar{C}),$$

$$\text{col}(C) \subset \text{col}(\bar{C}),$$

$$K_0 \leq \bar{K}_0.$$

It suffices to show that $x\bar{C} = 0$ implies $xC = 0$. If $x\bar{C} = 0$, then

$$A \text{diag}(xC)B' = \bar{A} \text{diag}(x\bar{C})\bar{B}' = 0,$$

so $\text{rank}[A \text{diag}(xC)B'] = 0$. By a calculation of rank like that done in the

Rank Lemma earlier in the paper, we have

$$0 = \text{rank} [A \text{diag}(xC) B'] \geq H_{AB}(\text{rank} \text{diag}(xC)).$$

By the assumed inequalities this can only happen if $\text{rank} \text{diag}(xC) = 0$, so $xC = 0$.

Let $w(y)$ indicate the number of nonzero entries in y . If $w(x\bar{C}) \leq R - \bar{K}_0 + 1$, then we claim that $w(xC) \leq w(x\bar{C})$. For we have

$$\begin{aligned} R - K_0 + 1 &\geq R - \bar{K}_0 + 1 \geq w(x\bar{C}) = \text{rank} \text{diag}(x\bar{C}) \\ &\geq \text{rank} [\bar{A} \text{diag}(x\bar{C}) \bar{B}'] = \text{rank} [A \text{diag}(xC) B'] \\ &\geq H_{AB}(\text{rank} \text{diag}(xC)) = H_{AB}(w(xC)). \end{aligned}$$

By the assumption on H_{AB} in Theorem 4d this implies $w(xC) \leq R - K_0 + 1$; hence by using the assumption again,

$$H_{AB}(w(xC)) = w(xC),$$

and our claim is sustained.

Now we invoke the Permutation Lemma, which is proved in a separate section, and find that there is a permutation matrix P_C and a nonsingular diagonal matrix N (nu) such that

$$\bar{C} = CP_C N.$$

A symmetrical argument yields that

$$\bar{B} = BP_B M,$$

where P_B is a permutation matrix and M (mu) is nonsingular diagonal. (Due to the asymmetry of the assumptions we cannot use the same technique to get P_A and Λ .) Since $J_1 \geq 2$, $K_1 \geq 2$, the extra hypothesis of the Permutation Lemma is valid (that is, every pair of columns of C is linearly independent, and likewise for B), so P_B , M , P_C , N are all unique. We can conclude also that $\text{rank}(\bar{B}) = \text{rank}(B) = J_0$, $\text{rank}(\bar{C}) = \text{rank}(C) = K_0$, every J_1 columns of \bar{B} are independent, and every K_1 columns of \bar{C} are independent.

Now we wish to prove that $P_B = P_C$. Suppose π_B and π_C are the permutations which correspond to P_B and P_C , so that if $b_{\downarrow r}$ is the r th column of B ,

etc., then

$$\bar{b}_{\downarrow \pi_B(r)} = b_{\downarrow r} \mu_r,$$

$$\bar{C}_{\downarrow \pi_C(r)} = c_{\downarrow r} \nu_r.$$

We wish to prove that π_B and π_C are equal. For this purpose, we introduce new notation for sets of columns of B and C and for sets of indices. We shall use S and T to indicate subsets of $\{1, 2, \dots, R\}$, and $-S$ and $-T$ to indicate their complementary sets. We shall use $b_{\downarrow S}$ to indicate the set of columns with subscripts in S , and similarly for other matrices.

If, contrary to what we wish to prove, π_B and π_C are not equal, then there is a value r_0 such that $s_0 =_{\text{def}} \pi_B(r_0) \neq t_0 =_{\text{def}} \pi_C(r_0)$. Now we claim that there are sets S and T with these properties (see Fig. 3):

$$S \cap T \text{ is empty,}$$

$$s_0 \text{ in } S, \quad t_0 \text{ in } T,$$

$$\bar{b}_{\downarrow -S} \text{ is a } \bar{B}\text{-hyperplane,} \quad \bar{c}_{\downarrow -T} \text{ is a } \bar{C}\text{-hyperplane.}$$

A set of columns in a matrix is defined to be a *hyperplane* (or “copoint”) in combinatorial geometry and matroid theory if it has rank one less than the matrix and is maximal with respect to this property. To construct S and T , we first pick a hyperplane \tilde{C} in \bar{C} such that $\bar{c}_{\downarrow s_0}$ is in \tilde{C} but $\bar{c}_{\downarrow t_0}$ is not. (Since every J_1 columns of \bar{C} are independent, and $J_1 \geq 2$, this is possible.) Then we let T be the indices which do not correspond to \tilde{C} , so that $\bar{c}_{\downarrow -T} = \tilde{C}$. Because \tilde{C} is a \bar{C} -hyperplane it has rank $K_0 - 1$, so we have

$$s_0 \text{ not in } T, \quad t_0 \text{ in } T,$$

$$\bar{c}_{\downarrow -T} \text{ is a } \bar{C}\text{-hyperplane,}$$

$$\text{card}(\tilde{C}) \geq K_0 - 1, \quad \text{card}(T) \leq R - K_0 + 1 \leq J_1 - 1.$$

Now consider $\bar{b}_{\downarrow T}$. Since every J_1 columns of \bar{B} are independent, $\bar{b}_{\downarrow T}$ spans only itself in \bar{B} . Also, $\bar{b}_{\downarrow s_0}$ is not in $\bar{b}_{\downarrow T}$. Therefore there is a hyperplane \bar{B} in \bar{B} which contains $\bar{b}_{\downarrow T}$ but does not contain $\bar{b}_{\downarrow s_0}$. Let S be the indices which

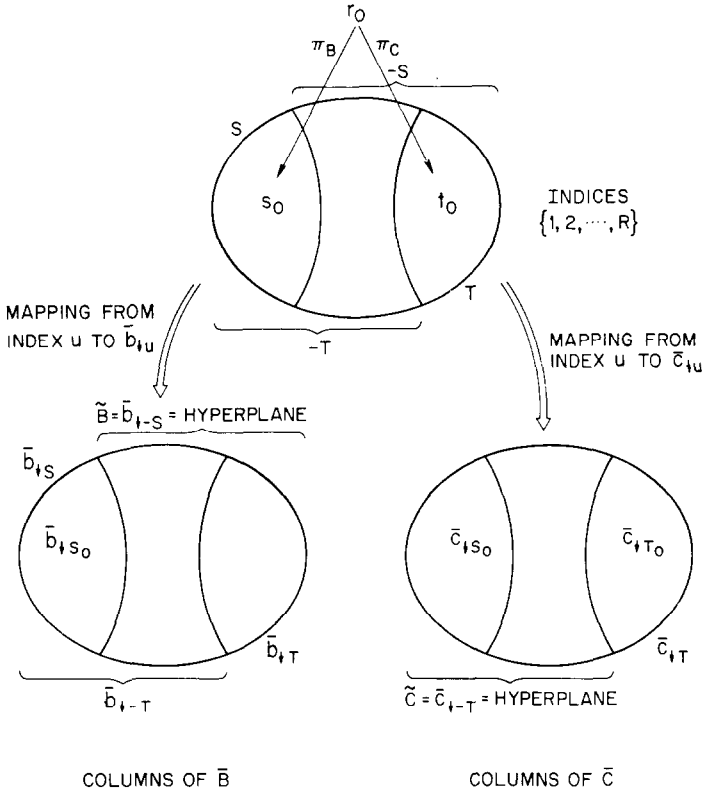


FIG. 3.

do not correspond to \tilde{B} , so that $b_{\downarrow -S} = \tilde{B}$. By construction

$$s_0 \text{ in } S, \quad (-S) \cup (-T) = \{1, 2, \dots, R\},$$

$\tilde{b}_{\downarrow -S}$ is a \tilde{B} -hyperplane.

It now follows that $S \cap T$ is empty, so S and T have all the properties claimed.

Now we use S and T to derive a contradiction. From the basic assumption of the theorem, we have for every v and w that

$$(vB) \text{diag}(a_{i\rightarrow}) (wC)' = (v\tilde{B}) \text{diag}(\bar{a}_{i\rightarrow}) (w\tilde{C})', \quad \text{all } i,$$

so

$$\begin{aligned} & A \operatorname{diag}(vB) \operatorname{diag}(wC) (1, \dots, 1)' \\ &= \bar{A} \operatorname{diag}(v\bar{B}) \operatorname{diag}(w\bar{C}) (1, \dots, 1)'. \end{aligned}$$

Since $\bar{b}_{\downarrow -S}$ is a hyperplane, we can pick v with the following property, and similarly we can pick w so that

$$\begin{aligned} (v\bar{B})_r & \begin{cases} \neq 0 & \text{for } r \text{ in } S, \\ = 0 & \text{for } r \text{ not in } S, \end{cases} \\ (w\bar{C})_r & \begin{cases} \neq 0 & \text{for } r \text{ in } T, \\ = 0 & \text{for } r \text{ not in } T. \end{cases} \end{aligned}$$

Since $S \cap T = \emptyset$, we see that

$$\operatorname{diag}(v\bar{B}) \operatorname{diag}(w\bar{C}) = 0.$$

so we get

$$A \operatorname{diag}(vB) \operatorname{diag}(wC) (1, \dots, 1)' = 0,$$

a linear relationship between the columns of A . Since

$$(vB)_r = vb_{\downarrow r} = \frac{1}{\mu_r} v \bar{b}_{\downarrow \pi_B(r)}, \quad (wC)_r = \frac{1}{\nu_r} w \bar{c}_{\downarrow \pi_C(r)},$$

$\pi_B(r_0)$ is in S , and $\pi_C(r_0)$ is in T , we see that the coefficient of the r_0 th column is nonzero, so this linear relationship is not trivial. At the same time, the number of columns of A with nonzero coefficients is

$$\operatorname{card}[\pi_B^{-1}(S) \cap \pi_C^{-1}(T)] \leq \min(\operatorname{card}(S), \operatorname{card}(T))$$

$$\leq \max(R - J_0 + 1, R - K_0 + 1) \leq I_1 - 1.$$

This contradicts the assumption that every I_1 columns of A are independent, and completes the proof that $P_B = P_C$.

Now that we have proved $P_B = P_C$, we shall write P for this common

permutation matrix, and we have

$$\bar{B} = BPM, \quad \bar{C} = CPN.$$

We now wish to prove that $\bar{A} = AP\Lambda$ for some diagonal matrix Λ . Note that

$$M \operatorname{diag}(y) N = \operatorname{diag}(yMN)$$

and

$$P \operatorname{diag}(y) P' = \operatorname{diag}(yP')$$

for any vector y . Therefore we have

$$B \operatorname{diag}(xA) C' = BPM \operatorname{diag}(x\bar{A}) NP' C' = B \operatorname{diag}(x\bar{A}MNP') C',$$

$$B \operatorname{diag}(x(A - \bar{A}MNP')) C' = 0,$$

$$0 = \operatorname{rank}[B \operatorname{diag}(\cdots) C'] \geq H_{BC}(\operatorname{rank} \operatorname{diag}(\cdots)) \geq 0$$

for all x . Therefore, by the assumption on H_{BC} in Theorem 4d, we must have $\operatorname{rank} \operatorname{diag}(\cdots) = 0$; hence $x(A - \bar{A}MNP') = 0$ for all x . Thus $A = \bar{A}MNP'$, so

$$\bar{A} = AP(MN)^{-1} = AP\Lambda$$

with $\Lambda = (MN)^{-1}$ diagonal and $\Lambda MN = \text{identity}$. This completes the proof of Theorem 4. ■

5. THE PERMUTATION LEMMA

Because the following lemma is of interest in itself, quite apart from its application in this paper, we state it and prove it without reference to the theorem or the other lemmas. It is helpful to think about this lemma geometrically in terms of the column vectors in A and in \bar{A} . We use $w(y)$ to indicate the number of nonzero components of y .

PERMUTATION LEMMA. *Suppose we are given two matrices A and \bar{A} , which are I by R and I by \bar{R} , where $R \geq \bar{R}$. Suppose A has no zero columns.*

Suppose that for any vector x such that

$$w(x\bar{A}) \leq \bar{R} - \text{rank}(\bar{A}) + 1,$$

we have

$$w(xA) \leq w(x\bar{A}).$$

Then $R = \bar{R}$, and there are a permutation matrix P_A and a nonsingular diagonal matrix Λ such that $\bar{A} = AP_A\Lambda$. In other words, A and \bar{A} have the same columns up to permutation and multiplication by nonzero scalars. Furthermore, if every two columns of A are linearly independent, then P_A and Λ are unique.

The most interesting case for this lemma occurs when $\text{rank}(A) < R$. Thus it is appropriate during the proof to think of A and \bar{A} as having more columns than rows.

Proof. Consider the columns of A and \bar{A} as vectors in I -dimensional column space. We reserve the phrase *i-dimensional subspace* to refer to any i -dimensional subspace (which contains the origin) of the entire I -dimensional column space. An *i-dimensional flat* F will mean a set of columns of A which is contained in some i -dimensional subspace, and which is maximal with respect to this property (that is, F is not contained in any larger set of columns which is contained in any i -dimensional subspace). Note that every i -flat contains at least i columns of \bar{A} . The 0-dimensional flat consists of all zero columns of \bar{A} (though this set will turn out to be empty).

Our main tool will be the following fact. For any i -dimensional flat, F , the $(i+1)$ -dimensional flats containing F partition the columns of \bar{A} not in F ; that is, each column of \bar{A} not in F belongs to precisely one $(i+1)$ -dimensional flat containing F . This is well known and can easily be verified directly. It is also one of the basic properties which is preserved when sets of vectors are generalized in combinatorial geometry and matroid theory.

Denote the rank of A and \bar{A} by I_0 and \bar{I}_0 . Let $\text{col}(A)$ be the subspace spanned by the columns of A , and let $\text{null}(A)$ be the set of all x such that $xA = 0$. Then $\text{col}(A)$ is an I_0 -dimensional subspace, and $\text{null}(A)$ is its $(I - I_0)$ -dimensional orthogonal complement. We make similar definitions and remarks for \bar{A} . For every x in $\text{null}(A)$, $w(xA) = 0$. By the fundamental assumption of the Permutation Lemma it then follows that $w(x\bar{A}) = 0$, so $x\bar{A} = 0$. It now easily follows that $\text{null}(A) \supset \text{null}(\bar{A})$, $\text{col}(A) \subset \text{col}(\bar{A})$, and $I_0 \leq \bar{I}_0$.

We shall now prove the following proposition, starting with large values of i and working downward by induction.

PROPOSITION. For any i -dimensional flat F , let S be the subspace it spans. Let $\bar{k} = \text{card}(F) =$ the number of columns of \bar{A} in S , and let $k =$ the number of columns of A in S . Then $k - R \geq \bar{k} - \bar{R}$.

The meaning of this inequality is simpler to see if we suppose $R = \bar{R}$: in this case it would say that for every subspace spanned by some flat, the subspace contains at least as many columns of A as of \bar{A} .

Proof of the Proposition. For the largest possible value of i , namely $i = \bar{I}_0$, there is only one i -dimensional flat, which contains all \bar{R} columns of \bar{A} , so $\bar{k} = \bar{R}$. We have proved that $\text{col}(A) \subset \text{col}(\bar{A})$, so $k = R$. As $R - R \geq \bar{R} - \bar{R}$, the proposition is true for $i = \bar{I}_0$.

Next consider $i = \bar{I}_0 - 1$. Let F be an i -dimensional flat, and let S be the subspace spanned by the columns in F . Then S is an $(\bar{I}_0 - 1)$ -dimensional subspace, so its orthogonal complement S^* contains $\text{null}(\bar{A})$ and has dimensionality one greater than that of $\text{null}(\bar{A})$. Let x be in $S^* - \text{null}(\bar{A})$, so $\{x, \text{null}(\bar{A})\}$ spans S^* . Then the zero coordinates of $x\bar{A}$ correspond to columns of \bar{A} in S , so $R - w(x\bar{A}) = \bar{k}$, and similarly $\bar{R} - w(xA) = k$. An i -dimensional flat contains at least i columns, so we have $\bar{k} \geq \bar{I}_0 - 1$, and

$$w(x\bar{A}) = \bar{R} - \bar{k} \leq \bar{R} - \bar{I}_0 + 1.$$

Then by the fundamental assumption of the Permutation Lemma, it follows that $w(xA) \leq w(x\bar{A})$, so

$$k - R = -w(xA) \geq -w(x\bar{A}) = \bar{k} - \bar{R},$$

which demonstrates the proposition for the case $i = \bar{I}_0 - 1$.

We know that the proposition holds for $i = \bar{I}_0$ and for $i = \bar{I}_0 - 1$. We proceed by induction, going downwards to smaller and smaller values of i . Suppose the proposition is already known to hold for $i + 1$, and we wish to prove it for i . Let F be an i -dimensional flat, and let $\{F_m\}$ with $m = 1$ to M be all the $(i + 1)$ -dimensional flats which contain F . Let S and $\{S_m\}$ be the subspaces generated by these flats. Let k and k_m be the number of columns of A contained in the same subspaces.

By a basic fact pointed out above, we know that the flats $\{F_m\}$ partition the columns of A which are not in F . Thus F and the sets $\{F_m - F\}$ partition

the \bar{R} columns of \bar{A} . Clearly $F_m - F$ contains $\bar{k}_m - \bar{k}$ columns, so we have

$$\bar{k} + \sum (\bar{k}_m - \bar{k}) = \bar{R},$$

and

$$\bar{k} - \bar{R} = \sum \frac{\bar{k}_m - R}{M-1}.$$

Now we use the induction hypothesis. Since F_m is an $(i+1)$ -dimensional flat, $k_m - R \geq \bar{k}_m - \bar{R}$. We do not yet know that $\cup S_m$ contains all the columns of A . However, whatever columns of A it does contain are partitioned by S and the sets $\{S_m - S\}$. These contain k and $\{k_m - k\}$ columns of A , so $\cup S_m$ contains

$$k + \sum (k_m - k) \leq R$$

columns of A . Then we have

$$\begin{aligned} k - R &\geq \sum \frac{k_m - R}{M-1} \\ &\geq \sum \frac{\bar{k}_m - \bar{R}}{M-1} = \bar{k} - \bar{R}, \end{aligned}$$

which completes the proof of the proposition. ■

Now we apply the proposition in the case $i=0$. The 0-dimensional flat consists of the zero columns of A . Thus \bar{k} = the number of zero columns of \bar{A} , and k = the number of zero columns of A . By assumption, $k=0$ and $R \geq \bar{R}$. By the proposition, $0 - R \geq \bar{k} - \bar{R}$, so

$$0 \geq \bar{R} - R \geq \bar{k} \geq 0,$$

which gives us $R = \bar{R}$ and $\bar{k} = 0$, so \bar{A} has no zero columns.

Consider the 1-dimensional flats F_m , and the subspaces S_m generated by these flats. Each F_m can be constructed by taking some column of \bar{A} and adjoining it to all other columns (if any) which are multiples of it. Clearly S_m consists of all possible multiples of some nonzero column in F_m . Let \bar{k}_m be the number of columns of \bar{A} contained in S_m , that is, $\bar{k}_m = \text{card}(F_m)$, and let

k_m be the number of columns of A contained in S_m . We claim that

$$R \geq \sum k_m \geq \sum \bar{k}_m \geq \bar{R} = R.$$

The first inequality follows because the S_m partition whatever columns of A are in $\cup S_m$ (this relies on the assumption that A contains no zero columns). Using $R = \bar{R}$ and the proposition for $i = 1$ gives $\bar{k}_m \geq \bar{k}_m$. The third inequality follows because $\cup S_m$ contains all columns of A . We now see that $k_m = \bar{k}_m$ for all m , and that $\cup S_m$ contains all columns of A . Therefore for every column of A , A and \bar{A} each contain exactly the same number of multiples of it.

Now we can choose a permutation π_A such that the $\pi(r)$ th column of \bar{A} is a nonzero multiple of the r th column of A . Furthermore, if every two columns of A are linearly independent, then $k_m = \bar{k}_m = 1$ for all m , and this permutation is unique. If P_A is a permutation matrix which accomplishes the permutation $\pi_{A, \bar{A}}$ and Λ is a diagonal matrix whose entries are the multipliers involved, then $\bar{A} = AP_A\Lambda$. ■

I would like to acknowledge the great stimulation provided by the paper [4] of Brockett and Dobkin, and my gratitude to Professor Roger Brockett for providing me with two versions of it. I would also like to thank Dr. Hans Witsenhausen for bringing their work to my attention.

REFERENCES

- 1 A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *The Design and Analysis of Computer Algorithms*, Addison-Wesley, Reading, Mass., 1974.
- 2 George M. Bergman, Ranks of tensors and change of base field, *J. Algebra* **11**, (1969) 613–621.
- 3 A. Borodin and I. Monroe, *Algebraic Complexity Theory*, American Elsevier, 1976.
- 4 Roger W. Brockett and David Dobkin, On the optimal evaluation of a set of bilinear forms, unpublished.
- 5 Roger W. Brockett and David Dobkin, On the number of multiplications required for matrix multiplication, unpublished.
- 6 J. D. Carroll and J. J. Chang, Analysis of individual differences in multidimensional scaling via an n -way generalization of Eckart-Yong decomposition, *Psychometrika* **35**, (1970) 283–319.
- 7 H. H. Crapo and Gian-Carlo Rota, *On the Foundations of Combinatorial Theory: Combinatorial Geometries* (prelim. ed.), M.I.T. Press, 1972.
- 8 Hans F. deGroote, On the complexity of quaternion-multiplication, *Inf. Process. Lett.* **3** (1975), 177–179.
- 9 C. M. Fiduccia and Y. Zalcstein, Algebras having linear multiplicative complexities, *J. Assoc. Comput. Mach.*, April, 1977.

- 10 F. G. Frobenius, Über den Rang einer Matrix, *Sitzungsber. K. Preuss. Akad. Wiss. Berl.* 1911, pp. 20–29, 128–129, reprinted in F. G. Frobenius, *Gesammelte Abhandlungen* (J-P. Serre, Ed.) Springer, Berlin, 1968, Paper 85, Vol. III, pp. 479–490.
- 11 T. D. Howell and Jean-Claude Lafon, The complexity of the quaternion product, Tech. Rep. 75–245, Dept. of Comput. Sci., Cornell Univ., 1975.
- 12 T. D. Howell, Tensor rank and the complexity of bilinear forms, Ph.D. thesis, Dept. of Comput. Sci., Cornell Univ., 1976.
- 13 J. B. Kruskal, More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling, *Psychometrika* **41** (1976), 281–293.
- 14 R. E. Miller and J. W. Thatcher (Eds.), *Complexity of Computer Computations*, Plenum, New York, 1972.
- 15 L. Mirsky, *An Introduction to Linear Algebra*, Oxford U. P., 1955.
- 16 Volker Strassen, Gaussian elimination is not optimal, *Numer. Math.* **13**, (1968), 354–356.
- 17 Volker Strassen, Some results in algebraic complexity theory, in *Proceedings of the International Congress of Mathematicians*, Vancouver, 1974.
- 18 Volker Strassen, Berechnung und Programm. I, *Acta Inf.* **1** (1971) 320–335.
- 19 Volker Strassen, Vermeidung von Divisionen, *J. Reine Angew. Math.* **264** (1973), 184–202.
- 20 J. van Leeuwen and P. van Emde Boas, Some elementary proofs of lower-bounds in complexity theory, unpublished report of the Stichting Mathematisch Centrum (Amsterdam), 1st ed., 1975; 2nd ed., 1976.

Received December 12, 1975