Bounds on the Ranks of Some 3-Tensors

M. D. Atkinson and S. Lloyd
Department of Computing Mathematics
University College
Cardiff, United Kingdom

Submitted by Richard A. Brualdi

ABSTRACT

The problem of obtaining upper bounds on the ranks of third order tensors is studied. New bounds are found using matrix canonical forms and nonlinear techniques from commutative algebra.

This paper addresses a problem originating in computer science which has attracted much recent attention [4, 8, 11, 12]. In computing terms the problem is to evaluate a number of bilinear forms in noncommuting variables using the smallest number of nonscalar multiplications (multiplications in which neither factor is a constant). Many multiplication problems in algebra are instances of this general problem; for example, forming matrix products [13], polynomial products [10], and group algebra products [1] all require several bilinear forms to be evaluated, and efficient algorithms for these problems are based on methods which use a small number of nonscalar multiplications.

The problem can also be posed in purely algebraic terms as follows: given a trilinear form $\sum_{i,j,k}\alpha_{ijk}x_iy_jz_k$ in variables $x_1,\dots,x_m,y_1,\dots,y_n,z_1,\dots,z_p$, find a representation of it as

$$\sum_{r=1}^{N} a_r(x)b_r(y)c_r(z) \qquad (*)$$

with a_r,b_r,c_r linear forms in x,y,z respectively and with N minimal. This minimal N is called the rank of the $m\times n\times p$ tensor (α_{ijk}). An equivalent way of defining the rank of (α_{ijk}) is in terms of the p matrices A_k whose (i,j) entry is α_{ijk}: the rank is the minimal number of rank 1 matrices whose linear span contains $A_1,\dots,A_p$. A full discussion of these ideas and much background information may be found in [12, 14].

0024-3795/80/030019+13$01.75

One can of course express a trilinear form $\sum \alpha_{ijk} x_i y_j z_k$ as $\sum_{i,j} x_i y_j \sum_k \alpha_{ijk} z_k$, which is a representation in the form (*) with $N = mn$. Thus the rank of any $m \times n \times p$ tensor cannot exceed mn (nor, similarly, np and pm). A fascinating aspect of this topic is that such trivial bounds can usually be substantially improved. An example of this, which also illustrates the problem in concrete terms, is provided by the following trilinear form:

$$x_1 y_2 z_3 + x_2 y_1 z_3 + x_1 y_3 z_2 + x_3 y_1 z_2 + x_2 y_3 z_1 + x_3 y_2 z_1.$$

Here it is already obvious that the rank is at most 6, but since the trilinear form can be written as

$$2XYZ + 2 \sum_{r=1}^{3} (x_r - X)(y_r - Y)(z_r - Z),$$

where $X = \frac{1}{2}(x_1 + x_2 + x_3)$, $Y = \frac{1}{2}(y_1 + y_2 + y_3)$, $Z = \frac{1}{2}(z_1 + z_2 + z_3)$, its rank is actually at most 4; in fact a short argument shows that its rank is exactly 4.

It is possible to extend the definition of rank to arbitrary t-tensors $(\alpha_{ij\ldots})$ with t subscripts. For $t = 2$ the rank of (α_{ij}) is just the ordinary matrix rank and so this case is fairly trivial, whilst progress for $t > 3$ seems to require a thorough understanding of third order tensors [5]. For this reason we restrict ourselves to third order tensors.

No general methods are known for calculating the rank of an arbitrary tensor; indeed, even the rank of the much-studied tensor which describes matrix multiplication is unknown. One concept which has proved helpful is equivalence. Two $m \times n \times p$ tensors $(\alpha_{ijk}), (\beta_{ijk})$ are said to be equivalent if

$$\beta_{ijk} = \sum_{u,v,w} r_{iu} s_{jv} t_{kw} \alpha_{uvw}$$

for some $m \times m, n \times n, p \times p$ nonsingular matrices $(r_{iu}), (s_{jv}), (t_{kw})$. If the $m \times n$ matrices $A_1, \ldots, A_p$ define an $m \times n \times p$ tensor, then all tensors equivalent to it can be obtained by first replacing $A_1, \ldots, A_p$ with another spanning set for $\langle A_1, \ldots, A_p \rangle$ and then performing row and column operations simultaneously on these new matrices.

Since equivalent tensors have the same rank, one is justified, when trying to compute the rank, in replacing a tensor by one equivalent to it. This simple fact together with canonical forms due to Kronecker allowed Grigoryev [7] and Ja'Ja' [9] independently to determine the exact rank of any $m \times n \times 2$ tensor in terms of certain invariants of its equivalence class. The lack of corresponding canonical forms for $p \geqslant 3$ indicates that a similar method will not work for a general $m \times n \times p$ tensor; indeed, it seems unlikely that a simple recipe for determining the rank of an arbitrary tensor exists. Because of this we shall study a more tractable problem, that of obtaining bounds on the rank valid for large classes of tensors. In this respect the paper is a sequel to [2], although we have tried to make our presentation self-contained. We here extend the techniques introduced in [2] and use them to obtain more refined upper bounds. These techniques differ from those previously used to obtain lower and upper bounds (for example, [4, 5, 11, 12]) in that they are highly nonlinear; we firmly believe that nonlinear analysis will be necessary to get the best results.

We define $r(m,n,p)$ to be the maximum rank, over the complex field C, achievable by an $m\times n\times p$ tensor (because the rank is dependent upon the underlying field of scalars, this field has to be specified; however, many of our techniques remain valid for arbitrary fields of characteristic zero). Some results about $r(m,n,p)$ can be found in [2, 5, 8]. In particular, it follows trivially from remarks above that $r(m,n,p)\leqslant \min(mn,np,pm)$. One of the main results of [2] was that $r(m,n,p)\leqslant m+\lfloor p/2\rfloor n$ if $m\leqslant n$. For $m=n$ this becomes

$$r(n,n,p)\leqslant\begin{cases}(p+1)n/2 & \text{if } p \text{ is odd,}\\ (p+2)n/2 & \text{if } p \text{ is even.}\end{cases}$$

Our first result is an improvement of the p even case.

THEOREM 1. $r(n,n,p)\leqslant (p+1)n/2$.

The improvement of this theorem is most noticeable for small p. Since the case $p=2$ has been settled completely by Grigoryev and Ja'Ja', the first case where an improvement is obtained is $p=4$, and we pause to make some remarks about this case. Here the theorem gives $r(n,n,4)\leqslant 5n/2$ rather than $r(n,n,4)\leqslant 3n$. Howell [8] and Brockett [3] independently showed that $r(m,n,p)\geqslant mnp/(m+n+p-2)$, and from this one can deduce that $r(n,n,4)\geqslant 2n-1$. In fact the exact values of $r(n,n,4)$ for $n=1,2,3$ are $1,4,6$ respectively, and it is known (Lloyd, unpublished) that $8\leqslant r(4,4,4)\leqslant 9$. Although we have not been able to tighten the lower and upper bounds for $r(n,n,4)$ in general, we have been able to give a criterion for the rank of an $n\times n\times 4$ tensor to be at most $2n$; this criterion almost always holds and, moreover, is easy to check. The criterion is the case $p=4$ of the next result.

THEOREM 2. *Let $A_1,\dots,A_p$ be $n\times n$ matrices defining an $n\times n\times p$ tensor (α_{ijk}), and let $x_1,\dots,x_p$ be indeterminates. Suppose that the determinant $|\Sigma x_iA_i|$, as a polynomial in $C[x_1,\dots,x_p]$, is not identically zero and has no repeated polynomial factor. Then the rank of (α_{ijk}) is at most $\lceil p/2\rceil n$.*

As a consequence of this and [3] one can deduce that almost all $n\times n\times 4$ tensors have rank $2n-1$ or $2n$. Before giving the proofs of Theorems 1 and 2 we recall that the discriminant of a polynomial $f(\lambda)=a_n\lambda^n+\cdots+a_0$ is a polynomial in the coefficients $a_0, a_1, \dots, a_n$ which vanishes if and only if $a_n=0$ or $f(\lambda)$ has a repeated factor [15]. We shall use this often; indeed, most of our arguments will be directed towards showing that certain polynomials do not have repeated factors. The first instance of this approach is

LEMMA A. *Let $f(x_1,\dots,x_m,y_1,\dots,y_n)$ be a polynomial in $m+n$ indeterminates with the property that, for all values of $y_1,\dots,y_n$, f has a repeated factor (as a polynomial in $C[x_1,\dots,x_m]$). Then f has a repeated factor as a polynomial in $C[x_1,\dots,x_m,y_1,\dots,y_n]$.*

Proof. We may regard f as a polynomial in any one of the x_i having coefficients in the field of rational functions of all the remaining $m+n-1$ variables, and its discriminant d_i is then a polynomial in these variables. For any values of $y_1,\dots,y_n$ we have $f=g^2h$ with g,h in $C[x_1,\dots,x_m]$ and g involving at least one of the x_i. Hence $d_i=0$ for some i. Therefore $d_1d_2\cdots d_m=0$ identically in $x_1,\dots,x_m,y_1,\dots,y_n$, and so some d_i is identically zero.

There is therefore some variable, x_1 say, for which we have $f=g^2h$ with g,h in $C(x_2,\dots,x_m,y_1,\dots,y_n)[x_1]$. But then a simple application of Gauss's lemma on primitive polynomials shows that g,h may be taken to be polynomials in $x_1,\dots,x_m,y_1,\dots,y_n$. ■

The next lemma follows from the proof of Theorem 1 of [2].

LEMMA B. *Let $A,B,C\dots$ be p $n\times n$ matrices defining an $n\times n\times p$ tensor (α_{ijk}), and suppose that A,B are simultaneously equivalent (by row and column operations) to diagonal matrices. Then the rank of (α_{ijk}) is at most $\lceil p/2\rceil n$.*

Proof of Theorem 2. Let $A_1,\dots,A_p$ satisfy the conditions of the theorem. Then, in particular, $|\Sigma x_iA_i|$ is not identically zero, and so $\langle A_1,\dots,A_p\rangle$ contains a nonsingular matrix. We may therefore replace the A_i's by another p-element generating set $B_1,\dots,B_p$ for $\langle A_1,\dots,A_p\rangle$ and have B_1 nonsingular. Since $|\Sigma x_iB_i|$ is obtained from $|\Sigma x_iA_i|$ by a nonsingular linear change of variables, it too has no repeated polynomial factor. Furthermore we may reduce B_1 to the identity matrix by row and column operations performed on each of $B_1,\dots,B_p$ giving matrices $C_1=I, C_2,\dots,C_p$, and again $|\Sigma x_iC_i|$ has no repeated polynomial factor.

From Lemma A, there exist values for $x_2,\ldots,x_p$ such that if $C=-\sum_{i=2}^{p}x_iC_i$, then $|x_1I-C|$ has distinct roots as a polynomial in x_1; in other words, C has distinct eigenvalues and so may be diagonalized. The original tensor is therefore equivalent to one defined by two diagonal matrices and $p-2$ further matrices, and so the theorem follows from Lemma B. ∎

LEMMA C. *If A,B are two $n\times n$ matrices, A is nonsingular, and the minimal rank of the nonzero matrices in $\langle A,B\rangle$ is r, then there exist $n-r$ rank 1 matrices $M_1,M_2,\ldots$ and constants $\beta_1,\beta_2,\ldots$ such that A and $B+\sum\beta_iM_i$ are simultaneously equivalent to diagonal matrices.*

Proof. We obviously lose no generality by replacing A,B with two equivalent matrices which, since A is nonsingular, may be taken as I,C. Similarly, if C has p invariant polynomials, then we may take C in rational canonical form having p companion matrices $C_1,\ldots,C_p$ on its diagonal, each one of them having minimal polynomial equal to an invariant polynomial. Since each invariant polynomial divides the next, the C_i have some common eigenvalue α. It follows that $\mathrm{rank}(C-\alpha I)\leqslant n-p$. If $C-\alpha I=0$, the lemma follows trivially, and so we shall assume that $C-\alpha I\neq 0$, in which case $r\leqslant n-p$, i.e. $p\leqslant n-r$.

To complete the proof we just have to show that given a companion matrix

$$C_i=\begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & 0 & a_1 \\ 1 & 0 & & & & & a_2 \\ & 1 & \cdot & & & & a_3 \\ & & \cdot & \cdot & & & \cdot \\ & & & \cdot & \cdot & & \cdot \\ & & & & \cdot & 0 & \cdot \\ & & & & & 1 & a_t \end{bmatrix},$$

there exists a rank 1 matrix M_i such that C_i+M_i is diagonalizable. To do this we take M_i to have nonzero entries in its last column only, in such a way that C_i+M_i is a companion matrix with distinct eigenvalues (recalling that the last column of a companion matrix gives the coefficients of its characteristic polynomial except for the leading coefficient). ∎

DEFINITION. A pair of $n\times n$ matrices A,B is said to be exactly deficient if the linear space $\langle A,B\rangle$ contains, except for the zero matrix, only matrices of rank $n-1$.

LEMMA D. *If A, B is an exactly deficient pair of $n \times n$ matrices, then there exists a rank 1 matrix M and constants α, β such that $A + \alpha M, B + \beta M$ are simultaneously equivalent to diagonal matrices.*

Proof. We may take A, B to be linearly independent; otherwise the lemma is trivial. Let $\lambda, \alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n$ be indeterminates, and let M be the $n \times n$ rank 1 matrix whose (i,j) entry is $x_i y_j$. Thus M is the "general rank 1 matrix" in that any particular rank 1 matrix can be obtained by giving appropriate values to $x_1, \dots, x_n, y_1, \dots, y_n$. If C is any matrix of rank $n-1$, there exists a rank 1 matrix D such that $|C + D| \neq 0$ and hence $|C + M| \not\equiv 0$; we use this fact several times.

Consider the polynomial $f(\lambda, \alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n) = |\lambda(A + \alpha M) - (B + \beta M)|$. If we can find values for $\alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n$ such that $|A + \alpha M| \neq 0$ and such that f becomes a polynomial in λ with distinct roots, then it is easy to prove that $A + \alpha M, B + \beta M$ are simultaneously diagonalizable. We shall therefore assume that for all values of $\alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n$ we have $|A + \alpha M| = 0$ or $d = 0$, where d is the discriminant of f with respect to λ. Consequently $|A + \alpha M| d$ is identically zero. Now $|A + \alpha M|$ is not identically zero, since $\operatorname{rank}(A) = n - 1$, and so d is identically zero. Therefore, as in the proof of Lemma A, $f = p^2 q$, where p, q are in $C[\lambda, \alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n]$.

In f the variables $x_1, \dots, x_n, y_1, \dots, y_n$ occur to the first power only, and so $p = p(\lambda, \alpha, \beta)$ is independent of them. Suppose there existed values of λ, α, β such that $p(\lambda, \alpha, \beta) = 0$ and $\lambda\alpha - \beta \neq 0$; then we would have $|\lambda A - B + \theta M| = 0$, $\theta \neq 0$, for all rank 1 matrices M, implying $\operatorname{rank}(\lambda A - B) < n - 1$, a contradiction. Thus $p(\lambda, \alpha, \beta) = 0$ implies $\lambda\alpha - \beta = 0$, and so, by the *Nullstellensatz*, $p(\lambda, \alpha, \beta)$ divides $(\lambda\alpha - \beta)^s$ for some s. However, $\lambda\alpha - \beta$ is irreducible, and so we have $f = (\lambda\alpha - \beta)^2 r$ with r in $C[\lambda, \alpha, \beta, x_1, \dots, x_n, y_1, \dots, y_n]$. But, upon expanding the determinant, we see that the coefficient of β^i is a sum of multiples of i-rowed minors of M all of which are zero for $i \geqslant 2$, and therefore $|\lambda(A + \alpha M) - (B + \beta M)| \equiv 0$. This is impossible, for it would imply $|B + \beta M| \equiv 0$, contradicting $\operatorname{rank}(B) = n - 1$. ∎

LEMMA E. *If A, B are $n \times n$ matrices having the property that $\langle A, B \rangle$ contains no nonzero matrix of rank less than or equal to $\lfloor n/2 \rfloor$, then there exist $\lfloor n/2 \rfloor$ rank 1 matrices $M_1, M_2, \dots$ and constants $\alpha_1, \alpha_2, \dots, \beta_1, \beta_2, \dots$ such that $A + \sum \alpha_i M_i$ and $B + \sum \beta_i M_i$ are simultaneously equivalent to diagonal matrices.*

Proof. As in the proof of Lemma C, we lose no generality by replacing A, B by equivalent matrices, and in this case we take the pencil $\lambda A + B$ to be in its Kronecker canonical form. Neither (since we may replace A, B by any two matrices which also generate $\langle A,B\rangle$) do we lose generality by assuming that the regular part of the pencil contains no infinite elementary divisors. Employing temporarily the notation of [6], we note that each of the square pencils

$$\begin{pmatrix} L_\varepsilon & 0 \\ 0 & L_\eta^T \end{pmatrix}, \quad \begin{pmatrix} L_\varepsilon & 0 \end{pmatrix}, \quad \text{and their transposes}$$

are exactly deficient pencils. From this we can conclude that the pencil decomposes as a direct sum of an $m\times m$ regular pencil λA_0+B_0, exactly deficient $s_i\times s_i$ pencils λA_i+B_i, $i=1,2,\dots,k$, and a $t\times t$ zero pencil. In particular $m+\Sigma s_i+t=n$.

Let r be the minimal rank of the nonzero matrices in $\langle A_0,B_0\rangle$, so that, from the hypotheses of the lemma, $r+\Sigma(s_i-1)>n/2$. We apply Lemma C to the regular pencil λA_0+B_0 and Lemma D to each of the exactly deficient pencils λA_i+B_i. This gives $m-r+k$ rank 1 matrices $M_1,M_2,\dots$ and constants $\alpha_1,\alpha_2,\dots,\beta_1,\beta_2,\dots$ such that $A+\Sigma\alpha_iM_i$, $B+\Sigma\beta_iM_i$ are simultaneously equivalent to diagonal matrices. The proof is completed by noting that

$$m-r+k\leqslant m-r+k+t=n-\sum s_i-r+k=n-r-\sum(s_i-1)<n-\frac{n}{2}=\frac{n}{2}.$$

■

Proof of Theorem 1. We have to show that an arbitrary $n\times n\times p$ tensor has rank at most $(p+1)n/2$. As noted previously, we may take p to be even, and then the tensor can be described by $p/2$ pairs of $n\times n$ matrices $A_1,B_1,A_2,B_2,\dots$. We have to find $\lfloor(p+1)n/2\rfloor$ rank 1 matrices whose span contains these matrices. This is relatively easy if the space $\langle A_1,B_1, A_2,B_2,\dots\rangle$ contains a nonzero matrix C of rank $r\leqslant n/2$. For then we may replace the original matrices by p matrices one of which is C, take r rank 1 matrices whose sum is C, and take $\lfloor pn/2\rfloor$ rank 1 matrices whose span include the other $p-1$ matrices.

Otherwise we apply Lemma E to A_1,B_1 to obtain $\lfloor n/2\rfloor$ rank 1 matrices $M_1,M_2,\dots$ such that $A_1'=A_1+\Sigma\alpha_iM_i$, $B_1'=B_1+\Sigma\beta_iM_i$ are simultaneously equivalent to diagonal matrices. Lemma B now shows that the tensor defined by $A_1',B_1',A_2,B_2,\dots$ has rank at most $\lceil p/2\rceil n$, and so the original tensor has rank at most $(p+1)n/2$. ■

To motivate our next results we return to the bound mentioned above: $r(m,n,p)\leqslant m+\lfloor p/2\rfloor n$ if $m\leqslant n$, and note that this bound is useless if

$2m \leqslant n$, since it is then weaker than the trivial bound $r(m,n,p) \leqslant \min(mn,np,pm)$. Perhaps the simplest case where the bound provides only trivial information is that of $n\times 2n\times 3$ tensors [where it gives $r(n,2n,3)\leqslant 3n$], and so we now consider how better bounds can be obtained in this case.

To begin with we note that there exist $n\times n$ matrices A,B defining an $n\times n\times 2$ tensor of rank $\lfloor 3n/2\rfloor$ [7, 9], and it is then easy to prove that the $n\times 2n\times 3$ tensor defined by the $n\times 2n$ matrices $(A\ \ 0)$, $(B\ \ 0)$, $(0\ \ I)$ has rank $\lfloor 5n/2\rfloor$. From this we can conclude that $r(n,2n,3)\geqslant 5n/2$. In fact for $1\leqslant n\leqslant 4$ detailed calculations show that $r(n,2n,3)=\lfloor 5n/2\rfloor$, and we conjecture that equality holds for all n. Although we cannot prove this conjecture, we give some supporting evidence for it below.

Suppose that E,F,G are $n\times 2n$ matrices defining an $n\times 2n\times 3$ tensor (α_{ijk}), and let $\mathcal{X}$ be the space generated by the $2n\times 2n$ matrices

$$\begin{pmatrix} E\\ 0\end{pmatrix},\ \begin{pmatrix} F\\ 0\end{pmatrix},\ \begin{pmatrix} G\\ 0\end{pmatrix},\ \begin{pmatrix} 0\\ E\end{pmatrix},\ \begin{pmatrix} 0\\ F\end{pmatrix},\ \begin{pmatrix} 0\\ G\end{pmatrix}.$$

If $\mathcal{X}$ contains a nonsingular matrix, we shall say that (α_{ijk}) satisfies the nonsingularity condition. Clearly, most tensors will satisfy this condition. It is not hard to see that under the nonsingularity condition (α_{ijk}) is equivalent to a tensor defined by the $n\times 2n$ matrices $(I\ \ 0)$, $(0\ \ I)$, $(A\ \ B)$. We give two results concerning such tensors.

THEOREM 3. *An $n\times 2n\times 3$ tensor satisfying the nonsingularity condition almost always has rank exactly $2n$.*

Proof. Consider any tensor defined by the $n\times 2n$ matrices $(I\ \ 0)$, $(0\ \ I)$, $(A\ \ B)$. It is almost always the case that A,B are each diagonalizable. In this case both the $n\times n\times 2$ tensors defined by I,A and by I,B have rank n. It follows immediately that the original tensor has rank at most $2n$. But then a simple argument shows that it has rank exactly $2n$. ■

THEOREM 4. *Every $n\times 2n\times 3$ tensor satisfying the nonsingularity condition has rank at most $\lfloor 5n/2\rfloor$.*

The main step in the proof of this theorem is in establishing the following lemma.

LEMMA F. *Let A be any $n\times n$ matrix, and let $X=(x_{ij})$ be an $n\times\lfloor n/2\rfloor$ matrix of indeterminates. Then there exists an $\lfloor n/2\rfloor\times n$ matrix Y such that the matrix $A+XY$ (whose entries are in the field $C(x_{11},x_{12},\dots)$) has multiplicity-free invariant factors.*

Proof. Suppose the lemma is true for a particular matrix A, i.e. for some Y, $A+XY$ has multiplicity-free invariant factors. Then for any nonsingular matrix P with complex entries the invariant factors of $P^{-1}AP+P^{-1}XYP$ are also multiplicity-free. Let $P^{-1}X=X'=(x'_{ij})$ and $YP=Y'$. Using $X=PX'$, we can substitute for the indeterminates x_{ij} in terms of the indeterminates x'_{ij} and this will not affect whether the invariant factors of $P^{-1}AP+X'Y'$ are multiplicity-free. In other words, if the lemma is true for A, it is also true for $P^{-1}AP$, and so no generality is lost in taking A in its rational canonical form. Thus

$$A=\begin{bmatrix} \alpha &&&&&&& \\ & \alpha &&&&&& \\ && \ddots &&&&& \\ &&& \alpha &&&& \\ &&&& C_1 &&& \\ &&&&& C_2 && \\ &&&&&& \ddots & \\ &&&&&&& C_q \end{bmatrix},$$

where α occurs p times and where $C_1, C_2, \ldots, C_q$ are all nontrivial companion matrices; in particular $q \leqslant \lfloor n/2 \rfloor$.

Let j_r be the column number in A where the last column of C_r occurs, and let Y be the $\lfloor n/2 \rfloor \times n$ matrix defined by

$$Y_{i,j_i}=1, \qquad i=1,2,\ldots,q,$$

$$Y_{rs}=0, \qquad \text{otherwise.}$$

It is then clear that XY is an $n\times n$ matrix consisting of zeros except for columns $j_1, j_2, \ldots, j_q$, in which appear columns $1,2,\ldots,q$ of X.

To calculate the invariant factors of $A+XY$ we have to perform row and column operations on $A+XY-\lambda I$ (working over the polynomial ring $C(x_{11},x_{12},\ldots)[\lambda]$) to bring it to a diagonal form in which each diagonal element [a polynomial in λ with coefficients in $C(x_{11},x_{12},\ldots)$] divides the next; the invariant factors are those entries which depend properly on λ. Before proceeding to do this we make one small change of variables to slightly ease the exposition. The entries in the j_ith column of $A+XY-\lambda I$ are usually single indeterminates x_{ki}; the only exceptions are in those rows which are rows of C_i, and here the indeterminates have a constant added to them (and in one case $-\lambda$ also added). The change of variables we make simply causes the entries in the j_ith column to be all single indeterminates (except for the one case where $-\lambda$ is added); this does not affect whether the invariant factors are multiplicity-free.

We now give the sequence of row and column operations which is to be applied to $A+XY-\lambda I$. Rather than give a "snapshot" of the result after each operation, we rely on the reader to follow through the calculations, checking details where necessary. Consider each companion submatrix C_k in turn. To its first row (or, more precisely, the row of $A+XY-\lambda I$ containing the first row of C_k) we add λ times the second row, λ^2 times the third row, etc., and then permute the rows so that it is upper triangular with diagonal entries $1,1,\dots,1,\psi(\lambda)$; $\psi(\lambda)$ is a polynomial with leading coefficient -1 and whose other coefficients are indeterminates. Having done this for each companion submatrix we perform column operations on the whole matrix so that the "1" entries on the diagonal are the only entries which occur in their row. Next we rearrange rows and columns so that all the 1's occur in the first group of diagonal positions, followed by diagonal entries $\alpha-\lambda$. This brings the matrix to the form

$$\begin{bmatrix} I & 0 & 0 \\ 0 & (\alpha-\lambda)I & W \\ 0 & 0 & Z \end{bmatrix},$$

where W is a $p\times q$ matrix of indeterminates and the entries of Z are nonlinear polynomials in λ; the coefficients of these polynomials are indeterminates with the exception of the leading terms of the diagonal entries: these are -1. All the indeterminates which occur in W and Z are distinct.

From now on it is obviously sufficient to consider just the $(p+q)\times(p+q)$ submatrix

$$T=\begin{pmatrix} (\alpha-\lambda)I_p & W \\ 0 & Z \end{pmatrix},$$

and in doing so we distinguish between two cases according to the shape of W.

(i) $p\geqslant q$. Since $\operatorname{rank}(W)=q$, there exists a $p\times p$ nonsingular matrix G such that

$$GW=\begin{pmatrix} I_q \\ 0 \end{pmatrix},$$

and we may replace T by

$$U=\begin{pmatrix} G & C \\ 0 & I \end{pmatrix} T \begin{pmatrix} G^{-1} & 0 \\ 0 & I \end{pmatrix} = \begin{bmatrix} (\alpha-\lambda)I_q & 0 & I_q \\ 0 & (\alpha-\lambda)I_{p-q} & 0 \\ 0 & 0 & Z \end{bmatrix}.$$

Then we replace U by

$$\begin{bmatrix} I_q & 0 & 0 \\ 0 & I_{p-q} & 0 \\ -Z & 0 & I_q \end{bmatrix} U \begin{bmatrix} -I_q & 0 & 0 \\ 0 & I_{p-q} & 0 \\ (\alpha-\lambda)I_q & 0 & I_q \end{bmatrix} = \begin{bmatrix} 0 & 0 & I_q \\ 0 & (\alpha-\lambda)I_{p-q} & 0 \\ (\alpha-\lambda)Z & 0 & 0 \end{bmatrix},$$

and from this we see that there are $p-q$ invariant factors $\alpha-\lambda$ together with invariant factors obtained from $(\alpha-\lambda)Z$.

The invariant factors obtained from $(\alpha-\lambda)Z$ either are further polynomials $\alpha-\lambda$ or have the form $\alpha-\lambda$ multiplied by an invariant factor of Z. Since the product of the invariant factors of Z is the determinant $|Z|$, this case will be completed if we can show that $|Z|$ is a multiplicity-free polynomial in λ without any factor $\alpha-\lambda$. This however is immediate when we observe that values for the indeterminates maybe chosen so that Z becomes a diagonal matrix with diagonal entries which are mutually coprime multiplicity-free polynomials and which are not divisible by $\alpha-\lambda$.

(ii) $p \leqslant q$. In this case $\operatorname{rank}(W)=p$ and there exists a nonsingular $q\times q$ matrix H such that $WH=(I_p\ \ 0)$. This allows us to replace T by

$$V=T\begin{pmatrix} I_p & 0 \\ 0 & H \end{pmatrix} = \begin{pmatrix} (\alpha-\lambda)I_p & I_p & 0 \\ 0 & Z_0 & Z_1 \end{pmatrix},$$

where $(Z_0\ \ Z_1)=ZH$. Next we replace V by

$$\begin{pmatrix} I_p & 0 \\ -Z_0 & I_q \end{pmatrix} V \begin{bmatrix} I_p & 0 & 0 \\ -(\alpha-\lambda)I_p & I_p & 0 \\ 0 & 0 & I_{q-p} \end{bmatrix} = \begin{pmatrix} 0 & I_p & 0 \\ (\lambda-\alpha)Z_0 & 0 & Z_1 \end{pmatrix},$$

and it follows that we need only consider those invariant factors which arise from $((\lambda-\alpha)Z_0\ \ Z_1)$.

Now Z has the same invariant factors as $(Z_0\ \ Z_1)$, and under case (i) we noted that these are multiplicity-free and are not divisible by $\lambda-\alpha$. Hence to complete this case it is sufficient to prove that the invariant factors of $((\lambda-\alpha)Z_0\ \ Z_1)$ are those of $(Z_0\ \ Z_1)$ except that p of them (including, possibly, "trivial" invariant factors of degree zero) are multiplied by $\lambda-\alpha$.

To do this we recall that if $d_k(Z)$ denotes the highest common factor of all $k\times k$ minors of Z, then the invariant factors of Z are given by $d_k(Z)/d_{k-1}(Z)$, $k=1,2,\ldots$. We also observe that for any k with $1\leqslant k\leqslant q-p$ some $k\times k$ minor of Z_1 is not divisible by $\lambda-\alpha$; this is a consequence of $|(Z_0\ \ Z_1)|$ not being divisible by $\lambda-\alpha$ and the Laplace expansion of $|(Z_0\ \ Z_1)|$ in terms of $k\times k$ minors of Z_1. From this it follows that

$$d_i((\lambda-\alpha)Z_0\quad Z_1)=d_i(Z_0\quad Z_1),\qquad i=1,2,\ldots,q-p$$

$$d_i((\lambda-\alpha)Z_0\quad Z_1)=\lambda^{i-(q-p)}d_i(Z_0\quad Z_1),\qquad i=q-p+1,\ldots,q.$$

The invariant factors of $((\lambda-\alpha)Z_0\ \ Z_1)$ therefore have the property required of them. ■

Proof of Theorem 4. If we apply Lemma F to the $n\times n$ matrices A and B in turn, we obtain $\lfloor n/2\rfloor\times n$ matrices Y and Z such that the invariant factors of $A+XY$ and $B+XZ$ are multiplicity-free. it then follows that there exist nonsingular matrices P,Q with entries in $C(x_{11},x_{12},\ldots)$ such that

$$P^{-1}(A+XY)P=\begin{bmatrix}a_1&&\\&\ddots&\\&&a_n\end{bmatrix}\quad\text{and}\quad Q^{-1}(B+XZ)Q=\begin{bmatrix}b_1&&\\&\ddots&\\&&b_n\end{bmatrix}.$$

These equations remain valid for any substitution of values for the x_{ij}'s provided this results in no division by zero. If we choose such values, then $A+XY,B+XZ$ will be complex matrices with the property that the $n\times 2n\times 3$ tensor defined by the $n\times 2n$ matrices $(I\ \ 0)$, $(0\ \ I)$, $(A+XY\ \ B+XZ)$ has rank $2n$. However, $(A+XY\ \ B+XZ)=(A\ \ B)+X(Y\ \ Z)$, and since $\operatorname{rank}(X(Y\ \ Z))\leqslant\lfloor n/2\rfloor$, it follows that the tensor defined by $(I\ \ 0)$, $(0\ \ I)$, $(A\ \ B)$ has rank at most $\lfloor 5n/2\rfloor$. ■

Finally we observe that the considerations behind the last two theorems generalize with little more than verbal changes to $n\times kn\times(k+1)$ tensors. Most tensors of this type satisfy a generalized nonsingularity condition and are equivalent to a tensor defined by the $k+1$ $n\times kn$ matrices

$$(I\quad 0\quad 0\quad\cdots\quad 0),\ (0\quad I\quad 0\quad\cdots\quad 0),\ldots,$$
$$(0\quad 0\quad\cdots\quad I),\ (A_1\quad A_2\quad\cdots\quad A_k)$$

(here, all the submatrices shown are $n\times n$ matrices).

THEOREM 3′. *Almost all tensors defined by matrices of the above form have rank exactly kn.*

THEOREM 4′. *All tensors defined by matrices of the above form have rank at most* $(k+\frac{1}{2})n$.

The case $k=n-1$ of these theorems, which concerns $n\times n\times(n^2-n)$ tensors, provides supporting evidence for the conjecture, made in [2], that $r(n,n,n^2-n)=n^2-\lceil n/2\rceil$.

REFERENCES

1 M. D. Atkinson, The complexity of group algebra computations, *Theoret. Computer Sci.* 5:205–209 (1977).
2 M. D. Atkinson and N. M. Stephens, On the maximal multiplicative complexity of a family of bilinear forms, *Linear Algebra and Appl.* 27:1–8 (1979).
3 R. W. Brockett, On the generic degree of a 3-tensor, unpublished typescript, Harvard Univ., 1976.
4 R. W. Brockett and D. Dobkin, On the optimal evaluation of a set of bilinear forms, *Linear Algebra and Appl.* 19:207–235 (1978).
5 D. Dobkin, On the optimal evaluation of a set of n-linear forms, in *Conference Record, 14th Annual Symposium on Switching and Automata Theory, Iowa City*, 1973, pp. 92–102.
6 F. R. Gantmacher, *The Theory of Matrices*, Vol. 2, Chelsea, New York, 1959.
7 D. Yu. Grigoryev, Some new bounds on tensor rank, LOMI preprint E-2-1978, Leningrad, 1978.
8 T. D. Howell, Global properties of tensor rank, *Linear Algebra and Appl.* 22:9–23 (1978).
9 J. Ja'Ja', Optimal evaluation of pairs of bilinear forms, in *Proceedings of the 10th Annual ACM Symposium on Theory of Computing*, 1978, pp. 173–183.
10 D. E. Knuth, *Semi-numerical Algorithms*, Addison-Wesley, Reading, Mass., 1969.
11 J. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics, *Linear Algebra and Appl.* 18:95–138 (1977).
12 J. C. Lafon, Optimum computation of p bilinear forms, *Linear Algebra and Appl.* 10:225–240 (1975).
13 V. Strassen, Gaussian elimination is not optimal, *Numer. Math.* 13:354–356 (1969).
14 V. Strassen, Vermeidung von Divisionen, *J. Reine Angew. Math.* 264:184–202 (1973).
15 B. L. van der Waerden, *Modern Algebra*, Vol. 1, Frederick, New York, 1953.

Received 15 January 1979; revised 9 July 1979