P. HENRICI. *Applied and Computational Complex Analysis.*, Vol. 2, ix + 662 pp. John Wiley and Sons, New York, 1977.

THIS book is the second volume of a three-volume work, devoted to a wide range of topics in the complex domain (for the review of Vol. 1, see *Zh. vychisl. Mat. mat. Fiz.*, 17, 5, 1330, 1977) (English translation: *USSR Comp. Math. math. Phys.*, 17, 5, 232, 1977). The scientific and systematic arrangement of the author in expounding the topics of the construction of exact or approximate solutions of problems of complex analysis were formulated in the review of Vol. 1. The volume, beginning with Chapter 8, is mainly devoted to the theory of special functions, integral transformations, asymptotic methods and the theory of continued fractions. Chapter 8 "Infinite products" contains an exposition of the theory of such products, connected with the theory of numbers, the theory of entire functions, gamma functions and beta functions. Here also is expounded the theory of contour integrals of the Mellin-Barnes type, containing products of gamma functions. Chapter 9 "Ordinary differential equations" is devoted to the analytic theory of differential equations. The differential equations of Kummer, Bessel and other functions are considered as concrete examples. In Chapter 10 "Integral transformations" the general theory of integral transformations and the operational calculus is expounded. The theory of the Fourier integral, the discrete Laplace transformation and many of their applications (including those in the theory of numbers, the theory of partial differential equations etc.) is treated in detail. Chapter 11 "Asymptotic methods" contains the general methods of asymptotic theory and their application to the solution of differential equations, to the calculation of special functions and in other sections of applied analysis. In the last chapter, Chapter 12, "Continued fractions", a detailed exposition of the theory of continued fractions is given. It is extremely comprehensive and contains many methods for the expansion of analytic functions in continued fractions, and a discussion of their advantages over expansions in series and infinite products. As in the previous chapters, as particular realizations of these general expansions, the corresponding expansions are obtained for many special functions.

Of particular value are the computational methods based on the theory of continued fractions expounded here.

All the chapters end with problems of varying difficulty, methodical hints for the conduct of seminars on the topics discussed, and bibliographical references. At the end of the book there is an extensive bibliography (of eight pages) and the necessary indexes. The book is a valuable textbook for a wide range of mathematicians and specialists in other sciences. It is worth translating it into Russian.

M. K. Kerimov.
Translated by J. Berry.

U.S.S.R. Comput. Maths. Math. Phys. Vol. 19, pp. 1–20
Pergamon Press Ltd. 1980. Printed in Great Britain.
0041–5553/79/0601–0001$07.50/0

THE ALGEBRAIC COMPLEXITY OF COMPUTING A FAMILY OF BILINEAR FORMS*

D. YU. GRIGOR'EV

Leningrad

(*Received* 28 *August* 1978; *revised* 22 *February* 1979)

Some aspects of the theory of the algebraic complexity of computations are investigated, namely, the complexity of the computation of certain sets of bilinear forms from the point of view of the number of multiplications and divisions. The complexity of the computation of a pair of bilinear forms is characterized. A new, close to linear, estimate is obtained for the complexity of computing a product of polynomials over a finite field. A group of non-singular linear tensor-preserving transformations is described. The behaviour almost everywhere of the rank in tensor space is considered.

A popular problem in the theory of the complexity of computations is that concerning the computation of a set of bilinear forms [1]. Such problems include e.g. multiplication of matrices of general type, or multiplication of two polynomials etc. Without affecting the theory of complexity, we can confine ourselves to bilinear forms of non-commutative variables (see below, Proposition 2). This restriction gives technical advantages, and can also be useful when constructing recursive procedures similar to Strassen's algorithm [2] for fast matrix multiplication. As a model of computations we shall use non-branching computations (straight-line computations). This choice is justified inasmuch as when programming such computations, we can manage without conditions that use branching. In the theory of algebraic complexity, it is this model which is usually employed (see [1–4]); it is described in Section 1.

The estimation of the complexity of computing a set of bilinear forms can be reduced to the estimation of a tensor rank [4]. Let $V_1, \ldots, V_k$ be vector spaces over a field F, which hereafter will be called the basis field, and let $\tau \in V_1 \otimes_F \ldots \otimes_F V_k$. The rank of tensor τ over the field F is given by the equation

$$\mathrm{Rg}_F(\tau) = \min\left\{ N : \tau = \sum_{1 \leqslant i \leqslant N} v_1^i \otimes \ldots \otimes v_k^i, \text{ where } v_j^i \in V_j,\ 1 \leqslant j \leqslant k \right\}.$$

If the field H is an extension of the field F, the tensor τ can be regarded as an element of the tensor product $(V_1 \otimes_F H) \otimes_H \ldots \otimes_H (V_k \otimes_F H)$ and we can find $\mathrm{Rg}_H(\tau)$. Obviously, $\mathrm{Rg}_H(\tau) \leqslant \mathrm{Rg}_F(\tau)$.

This concept of tensor rank generalizes the usual concept of matrix rank, namely: if $\tau \in V_1^* \otimes_F V_2$ is a matrix, the $\mathrm{Rg}_F(A)$ is the same as the usual rank of the matrix A. The matrix rank is unaffected by extension of the basis field. We cannot say the same about the tensor rank, as will be seen later. Below, when it is clear what basis field we are talking about, no indication

Zh. vychisl. Mat. mat. Fiz., 19, 3, 563–580, 1979.

of it will be given. If $A^{(1)}, \ldots, A^{(p)}$ is a set of matrices, then the rank of the set (denoted by $\mathrm{Rg}(A^{(1)}, \ldots, A^{(p)})$) will be defined as the rank of the tensor τ, where τ_{ijk} is the (i, j) element of the matrix $A^{(k)}$.

It seems that the first systematic study of the concept of a tensor rank was undertaken in [4], then later in [5, 6]. Prior to this, characteristics similar to rank were studied e.g. in [3]. We consider rank in the present paper and obtain new properties and estimates.

We obtain new upper bounds for the complexity of computing a convolution (or a product of polynomials) over a finite field, and allied problems. When the basis field contains not more than $2n + 1$ elements, the complexity of multiplying polynomials of degree n is equal to $2n+1$ (see [6]; we recall that, in the present paper, complexity with respect to the number of binary multiplications and divisions is considered). In several papers, procedures have been described for multiplying n-ary numbers, from which are easily derived procedures with close estimates of complexity for computing a convolution (for a survey, see [7]). The first such procedure was described in [8] and gave an estimate $O(n^{\log_2 3})$. The use of the fast Fourier transformation [9] gives an estimate $O(n \ln n \ln \ln n)$ (see [10]) for the complexity of polynomial multiplication (over an arbitrary field). In Theorem 4 we obtain a new upper bound for the complexity (with respect to the number of multiplications and divisions) of computing a convolution over a finite field. It has the form $ng(n)$, where $g(n)$ increases with n, more slowly than any finite iteration of the logarithm. This estimate is not an improvement on the familiar estimate for the complexity with respect to the number of all operations [10].

1. Algebraic complexity of computations. The connection with tensor ranks

In this section, the basis field F is assumed fixed.

A non-branching computation (n.c.) of the input variables $x_1, \ldots, x_n$ is defined as a list of rows, the l-th of which has the form

$$t_l \leftarrow f(w_1, \ldots, w_k), \qquad 1 \leqslant l \leqslant m,$$

where t_l are distinct variables, different from $x_1, \ldots, x_n$; f is one of the following operations: addition, multiplication ($k = 2$), taking the inverse element ($k = 1$), multiplication by a fixed element of F ($k = 1$); every w_i is either some t_j, $j < l$, or one of the $x_1, \ldots, x_n$.

Every n.c. can be written as an acyclical flow graph. We choose $n + m$ nodes, numbered $1, 2, \ldots, n + m$. To the nodes with numbers $1, 2, \ldots, n$ we assign variables $x_1, \ldots, x_n$. We consider the line (1.1), and assume that variables $w_1, \ldots, w_k$ are already assigned to the nodes $c_1, \ldots, c_k$. Then, to node number $l + n$ we assign t_l, and draw the ribs from $c_1, \ldots, c_k$ to the node number $l + n$; we mark this node by a functional sign f. We define by induction on l for the value of the node (or of the variable assigned to it). The value of the variable x_i is the variable itself. The value of the variable t_l is the term $f(W_1, \ldots, W_k)$ of variables $x_1, \ldots, x_n$, where W_i is the value of the variable w_i, $1 \leqslant i \leqslant k$.

We fix the point $(\alpha_1, \ldots, \alpha_n) \in F^n$. If the following construction can be performed for some term, then as a result we associate with it a series of variables $x_1 - \alpha_1, \ldots, x_n - \alpha_n$ and then say that the term is defined at the point $(\alpha_1, \ldots, \alpha_n)$). The construction is performed by induction on the construction of the term. Let us confine ourselves to inversion. If term a is associated with series $(\beta - P)$, where β is its unattached term ($0 \neq \beta \in F$), then with the term a^{-1} we associate the series

$$\beta^{-1} \sum_{m \geqslant 0} (\beta^{-1} P)^m.$$

We shall say that two terms are equal in the set $S \subseteq F^n$ if, given any point $(\alpha_1, \ldots, \alpha_n) \in S$ the terms in it are defined and the corresponding coefficients of the series associated with the terms are equal. We shall say that the family of terms $h_1, \ldots, h_p$ of variables $x_1, \ldots, x_n$ is computed by the given n.c. in the set $S \subseteq F^n$ if the values of all variables of the given n.c. are defined at all points of S, and for every i, $1 \leqslant i \leqslant p$, the term h_i is equal to the value of one of the variables of the given n.c. in S.

When the field F is finite, the n.c. is a scheme of functional elements [11].

To each operation (see (1.1)) we assign a non-negative number (weight), and the complexity of the given n.c. is defined as the sum of the weights of the operations marking the nodes, with regard to all nodes of the relevant flow graph. Depending on the assignment of the weights, different versions of the concept of complexity are obtained. The algebraic complexity of the computation of the family of terms $h_1, \ldots, h_p$ of variables $x_1, \ldots, x_n$ in the set S is defined as the least complexity of the n.c. of variables $x_1, \ldots, x_n$, computing $h_1, \ldots, h_p$ in the set S. The usual (and basic) problem of the theory of computational complexity consists in estimating the complexity of computing a given family of terms in a given set. In the present paper we fix on the following concept of complexity. We assign to binary multiplication and inversion the weight 1, and to other operations the weight 0. The concept of complexity then arising has been assumed by many authors (see e.g. [1, 3, 4]).

The following proposition is due to Strassen (see [4], where he stated it for the case when the variables $x_1, \ldots, x_n$ commute, though his proof extends to the non-commutative case).

Proposition 1

If a family of elements of a free algebra of $x_1, \ldots, x_n$, the degree of each of which is not greater than d, is computed in a set S by some n.c. of complexity N, then we can construct an n.c. computing this family in S, of complexity not greater than $Nd(d-1)/2$ without inversions, and with degrees of values of all variables not greater than d.

Hence, when estimating the complexity of computing a family of elements of a free algebra, we can confine ourselves to n.c. with the properties in the conclusion of Proposition 1. If the computed elements are bilinear forms, and in some line a product of non-constants $\Omega = \Phi \cdot \Psi$ is computed, then Φ and Ψ are of degree 1, and Ω is of degree 2.

Proposition 2 (see [3])

Let $\mathscr{A}$ be a family of bilinear forms, containing only monomials of the type x_iy_j, and let the n.c. $\mathscr{C}$ of complexity N compute $\mathscr{A}$. Then:

(1) conditions of commutation of the variables $x_iy_j=y_jx_i$ for all i, j enable us at most to halve the complexity of computing (by means of n.c.) the family $\mathscr{A}$;

(2) when the variables do not commute, all the binary multiplications in the n.c. $\mathscr{C}$ can be replaced (without increasing the complexity N) by multiplications of the type $\Omega=\Phi.\Psi$, where Φ is a linear form of x_i, and Ψ a linear form of y_j.

Hence, when estimating the complexity of computing a family of bilinear forms, we can confine ourselves to an n.c. of the type indicated in Para. 2).

Proposition 3 (see [4])

Let $\{\mathscr{A}_1, \ldots, \mathscr{A}_k\}$ be a family of bilinear forms with matrices of coefficients $\{A_1, \ldots, A_k\}$ respectively. The following three numbers are identical:

(a) the algebraic complexity of computing the family $\{\mathscr{A}_1, \ldots, \mathscr{A}_k\}$;

(b) the rank $\mathrm{Rg}(A_1, \ldots, A_k)$;

(c) the least number N of matrices $B_1, \ldots, B_N$ of rank 1 such that all the $A_1, \ldots, A_k$ lie in their linear hull.

By using Propositions 1–3 we can reduce estimation of the algebraic complexity of computing a family of bilinear forms to estimation of the rank of a family of matrices (or a tensor rank).

2. Rank of a pair of matrices

We fix the basis field F. Let us examine the first non-trivial case of a rank, namely, the rank of a pair of matrices. In Lemmas 1 and 2, and Theorem 1 below, we assume that matrices A, B, C, C' are $n \times n$ (n is fixed). We can write any matrix A of rank 1 as a tensor product $A=u\otimes v$, where $u\in F^n$, $v\in(F^n)^*$. Notice also that, for any vector u and covector v, we have $\mathrm{Rg}(u\otimes v)=1$. We shall call the system of vectors $\{u_1, \ldots, u_k\}$ and of covectors $\{v_1, \ldots, v_k\}$ conjugate, if $(u_i, v_j)=\delta_{ij}$ (Kronecker delta).

We introduce the auxiliary relation $\preccurlyeq$ between pairs of matrices A, B:

$$B\preccurlyeq A \Leftrightarrow \mathrm{Rg}(A, B)=\mathrm{Rg}(A).$$

Lemma 1

Given any matrices A and B, the relation $B\preccurlyeq A$ is equivalent to the existence of a matrix C such that the following is satisfied: (1) $B=AC$; (2) C has a proper basis (we call such matrices diagonalizable); (3) $\mathrm{Ker}\, C\supseteq\mathrm{Ker}\, B\supseteq\mathrm{Ker}\, A$.

Proof. Let a matrix C that satisfies conditions (1)–(3) exist. By (2) and (3), vectors $u_1, \ldots, u_r$ exist, such that their linear hull $L(u_1, \ldots, u_r)$ is complementary to Ker A and $Cu_i=\lambda_iu_i$, $1\leqslant i\leqslant r$, for certain $\lambda_i\in F$, where $r=\mathrm{Rg}\,A$. We choose in subspace $(\mathrm{Ker}\,A)^\perp$ vectors $u_1', \ldots, u_r'$, conjugate to $u_1, \ldots, u_r$. Then,

$$A=\sum_{1\leqslant i\leqslant r}(Au_i)\otimes u_i', \qquad B=\sum_{1\leqslant i\leqslant r}\lambda_i(Au_i)\otimes u_i'. \tag{2.1}$$

It follows from (2.1) that $\mathrm{Rg}(A, B)\leqslant r=\mathrm{Rg}(A)$.

Conversely, let $\mathrm{Rg}(A, B)=\mathrm{Rg}(A)=r$. Then, for certain u_i, v_i, λ_i, $1\leqslant i\leqslant r$,

$$A=\sum_{1\leqslant i\leqslant r}u_i\otimes v_i, \qquad B=\sum_{1\leqslant i\leqslant r}\lambda_iu_i\otimes v_i, \tag{2.2}$$

where the covectors $v_1, \ldots, v_r$ are linearly independent (like vectors $u_1, \ldots, u_r$).

We choose a family of vectors $\{v_1', \ldots, v_r'\}$, conjugate to $\{v_1, \ldots, v_r\}$. We define the matrix C by the equations

$$Cv_i'=\lambda_iv_i', \quad 1\leqslant i\leqslant r, \quad C(\mathrm{Ker}\,A)=\{0\}. \tag{2.3}$$

The matrix C is correctly defined, since $L(v_1', \ldots, v_r')\cap\mathrm{Ker}\,A=\{0\}$. Let us check that C is the required matrix. In view of (2.3), it is diagonalizable. The inclusion $\mathrm{Ker}\,A\subseteq\mathrm{Ker}\,B$ follows from (2.2) and the linear independence of $u_1, \ldots, u_r$. From the inclusion $\mathrm{Ker}\,A\subseteq\mathrm{Ker}\,B$, and from (2.3), we have $B=AC$. To prove the inclusion $\mathrm{Ker}\,B\subseteq\mathrm{Ker}\,C$, we resolve an arbitrary vector $w\in\mathrm{Ker}\,B$ as $w=u+v$, where $u\in\mathrm{Ker}\,A$, $v\in L(v_1', \ldots, v_r')$. Then, $Bv=0$, and hence, from the equations $B=AC$ and $\mathrm{Ker}\,A\cap L(v_1', \ldots, v_r')=\{0\}$ we have $Cv=0$, i.e., by (2.3), $Cw=0$. The lemma is proved.

We define the conditional rank of matrix B with respect to matrix A:

$$\mathrm{Rg}(B|A)=\min_{C\ A}\mathrm{Rg}(B-C). \tag{2.4}$$

Theorem 1 (reductive formula for the rank of a pair of matrices)

$$\mathrm{Rg}(A, B)=\mathrm{Rg}(A)+\mathrm{Rg}(B|A).$$

The particular case of the theorem, when $A = E$ is the identity matrix, was proved in [9].

Proof. The inequality $\mathrm{Rg}(A, B) \leqslant \mathrm{Rg}\, A + \mathrm{Rg}(B|A)$ follows from (2.4).

Let us prove the reverse inequality. Let $\mathrm{Rg}(A, B) = r$. Then, vectors u_i, covectors v_i, and elements λ_i^1, λ_i^2 of the field F exist, such that

$$A = \sum_{1 \leqslant i \leqslant r} \lambda_i^1 u_i \otimes v_i, \qquad B = \sum_{1 \leqslant i \leqslant r} \lambda_i^2 u_i \otimes v_i.$$

Let the covectors $\{v_1, \ldots, v_l\}$ with some l, $0 \leqslant l \leqslant r$, be (after renumbering) the maximal linearly independent system of covectors among the v_i, $1 \leqslant i \leqslant r$, for i such that $\lambda_i^1 \neq 0$. We choose a family of vectors $\{v_1', \ldots, v_l'\}$, conjugate to $\{v_1, \ldots, v_l\}$. We define the matrix C' by the relations

$$C' v_i' = (\lambda_i^2 / \lambda_i^1) v_i', \quad 1 \leqslant i \leqslant l, \qquad C'((L(v_1, \ldots, v_l))^\perp) = \{0\}.$$

If $\lambda_m^1 \neq 0$ for some $m > l$, then $\gamma_i^m \in F$, $1 \leqslant i \leqslant l$, exist, such that

$$v_m = \sum_{1 \leqslant i \leqslant l} \gamma_i^m v_i.$$

The matrix AC' is then equal to

$$\sum_{1 \leqslant i \leqslant l} \lambda_i^2 u_i \otimes v_i + \sum_{l < m \leqslant r, \lambda_m^1 \neq 0} u_m \otimes \sum_{1 \leqslant i \leqslant l} (\lambda_i^2 / \lambda_i^1) \gamma_i^m v_i.$$

It follows from (2.5) that the matrix $B - AC'$ is equal to

$$\sum_{l < m \leqslant r, \lambda_m^1 \neq 0} u_m \otimes \left[\lambda_m^2 v_m - \sum_{1 \leqslant i \leqslant l} (\lambda_i^2 / \lambda_i^1) \gamma_i^m v_i \right] + \sum_{l < m \leqslant r, \lambda_m^1 = 0} u_m \otimes \lambda_m^2 v_m.$$

Hence $\mathrm{Rg}(B - AC') \leqslant r - l$.

Spaces $\mathrm{Ker}\, C'$ and $\mathrm{Ker}\, A$ have at the least a common part $(L(v_1, \ldots, v_l))^\perp$, whose dimensionality is $n - l$. We renumber $v'_1, \ldots, v'_l$ in such a way that, for some s, $0 \leqslant s \leqslant l$, the vectors $v'_1, \ldots, v'_s$ are the maximal system such that

$$L(v_1', \ldots, v_s') \cap \mathrm{Ker}\, A = \{0\}.$$

Let us show that $s = \mathrm{Rg}(A)$. It follows from (2.6) that $\mathrm{Rg}(A) \geqslant s$. On the other hand, $\dim \mathrm{Ker}\, A \geqslant \dim [(L(v_1, \ldots, v_l))^\perp] + \dim [L(v_1', \ldots, v_l') \cap \mathrm{Ker}\, A] = (n - l) + (l - s)$, and hence $\mathrm{Rg}\, A \leqslant s$. We define the matrix C by the equations

$$C v_i' = (\lambda_i^2 / \lambda_i^1) v_i', \quad 1 \leqslant i \leqslant s, \qquad C(\mathrm{Ker}\, A) = \{0\}.$$

The matrix C differs from C' in not more than $l - s$ vectors v_i', $s < i \leqslant l$, of its basis. Hence $\mathrm{Rg}(C - C') \leqslant l - s$. Hence

$$\mathrm{Rg}(B - AC) \leqslant \mathrm{Rg}(B - AC') + \mathrm{Rg}(A(C' - C)) \leqslant (r - l) + (l - s). \tag{2.8}$$

By (2.7), conditions (2) and (3) of Lemma 1 hold for the matrix C. Hence, by Lemma 1, we have $AC \leqslant A$. Hence, from the equation $s = \mathrm{Rg}(A)$, and from (2.8) and (2.4), we obtain

$$\mathrm{Rg}(A, B) = r \geqslant \mathrm{Rg}(A) + \mathrm{Rg}(B - AC) \geqslant \mathrm{Rg}(A) + \mathrm{Rg}(B|A),$$

which completes the proof of Theorem 1.

It will henceforth be assumed in this section that F is algebraically closed. As the topology below, we always take the Zariski topology [12]. The bar over a set denotes its closure.

Lemma 2

The set of matrix pairs A and B, for which $\mathrm{Rg}(A, B) \neq n$, lies in some closed set of dimensionality less than $2n^2$ (in the cases when the property holds in an open set, we shall say that it holds "almost everywhere".)

Proof. We fix two $n \times n$ matrices A and B. Let C be the associated matrix for A (see [13]). Let Δ be the discriminant of the characteristic polynomial of the matrix CB. We shall show that, if $\det A \neq 0$, then $\mathrm{Rg}(A, B) = n$. The lemma follows from this. Since $\det A \neq 0$, then $\mathrm{Rg}(A, B) = \mathrm{Rg}(E, A^{-1}B) = \mathrm{Rg}(E, CB)$, and since $\Delta \neq 0$ and all the roots of the characteristic polynomial of CB are distinct, we have $CB \leqslant E$ by Lemma 1. By Theorem 1, $\mathrm{Rg}(E, CB) = n + \mathrm{Rg}(CB \mid E) = n$ in accordance with (2.4).

Throughout what follows, let A and B be $m \times n$ matrices ($m \leqslant n$).

Theorem 2

The rank of a pair of rectangular $m \times n$ matrices ($m \leqslant n$) is almost everywhere equal to $\min(n, 2m)$.

Proof. We define a morphism φ_{mn} of the space M_{nn} of $n \times n$ matrices into the space M_{mn} of $m \times n$ matrices by putting $(\varphi_{mn} C)_{ij} = C_{ij}$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. We continue it uniquely to pairs of matrices and preserve its old notation. Let $\mathcal{U} \subseteq M_{nn} \times M_{nn}$ be the open set of Lemma 2, in which the rank is equal to n. Then, by Chevalley's theorem ([12], p. 37), equal to the union of differences of closed sets

$$\varphi_{mn}(\mathcal{U}) = \bigcup_i (\mathcal{Y}_i \setminus \mathcal{Y}'_i),$$

where $\mathcal{Y}_i$, $\mathcal{Y}'_i$ are closed. If we can show that $\overline{\varphi_{mn}(\mathcal{U})} = M_{mn} \times M_{mn}$, it will follow that $\mathcal{Y}_i = M_{mn} \times M_{mn}$ for some i, and then, $\varphi_{mn}(\mathcal{U})$ will contain an open set $(M_{mn} \times M_{mn}) \setminus \mathcal{Y}'_i$. Assume that this is not the case. Then, $\mathcal{Y} = (M_{mn} \times M_{mn}) \setminus \overline{\varphi_{mn}(\mathcal{U})}$ is a non-empty open set. Since φ_{mn} is a morphism, its preimage $\varphi_{mn}^{-1}(\mathcal{Y})$ is open, and since φ_{mn} is surjective, $\varphi_{mn}^{-1}(\mathcal{Y})$ is non-empty, and hence $\varphi_{mn}^{-1}(\mathcal{Y}) \cap \mathcal{U} \neq \varnothing$, which is a contradiction. Hence it follows that $\overline{\varphi_{mn}(\mathcal{U})} = M_{mn} \times M_{mn}$, i.e., the rank of a pair of $m \times n$ matrices is almost everywhere not greater than n, since $\mathrm{Rg}(\varphi_{mn}(C, C')) \leqslant \mathrm{Rg}(C, C') = n$, $(C, C') \in \mathcal{U}$. It is also obvious that, if $n \geqslant 2m$, then $\mathrm{Rg}(A, B) \leqslant \mathrm{Rg}(A) + \mathrm{Rg}(B) \leqslant 2m$.

Conversely, if $A_1, \ldots, A_m$ are vector rows of matrix A, and $B_1, \ldots, B_m$ are vector rows of B, then $\mathrm{Rg}(A_1, \ldots, A_m, B_1, \ldots, B_m)$ is almost everywhere equal to min and hence, almost everywhere, $\mathrm{Rg}(A, B) \geqslant \min\{n, 2m\}$. The theorem follows from these inequalities.

Let us turn to computing the rank of a matrix pair. Since $\mathrm{Rg}(A, B) = \mathrm{Rg}(DAC, DBC)$ where C and D are non-singular square matrices, then it is sufficient to evaluate the rank of the matrix pair in the Weierstrass–Kronecker canonical form (see [13], chapter 12). We shall use the terminology of [13] below.

Theorem 3

Let the matrix bundle $\lambda A + \mu B$ have minimal indices for columns $\{a_i\}$, and minimal indices for rows $\{b_j\}$. Let the regular "kernel" of the bundle be $p \times p$, and for every γ, let it contain d_γ elementary divisors of the type $(\alpha\lambda+\beta\mu)^s$ for $s \geqslant 2$ and $\alpha/\beta = \gamma$ (possibly $\gamma = \infty$). We put $d = \max_\gamma d_\gamma$. Then,

$$\mathrm{Rg}(A, B) = \sum_i (a_i+1) + \sum_j (b_j+1) + p + d.$$

The proof is based on Lemma 1 and Theorem 1 and may be found in [14]. The author recently found that a similar result is obtained in [15].

Corollary. We have $\max \mathrm{Rg}(A, B) = \min\{m + [n/2], 2m\}$, where $[x]$ is the integral part of x.

To prove that the estimate is reached in the case $[n/2] \leqslant m$, it is sufficient to take as $\lambda A + \mu B$ the bundle with parameters $a_1 = \ldots = a_{n-m} = 1$, $p = 2m-n$, $d = [p/2]$.

The next two notes show the difference in properties between the rank of a matrix pair and the rank of a single matrix.

Note 1. The rank of a single matrix is a lower semicontinuous function. By Theorems 2 and 3, this is not true for a rank of a matrix pair. It is easy to give an example with $n = 2$: let the matrix

$$A_{(\alpha)} = \begin{Vmatrix} 0 & 1 \\ \alpha & 0 \end{Vmatrix},$$

then $\mathrm{Rg}_C(E, A_{(\alpha)}) = 2$ for $\alpha \neq 0$ by Lemma 1 and Theorem 1, while $\mathrm{Rg}_C(E, A_{(0)}) = 3$ by Theorem 3.

Note 2. Consider the dependence of $\mathrm{Rg}_F(\tau)$ on the choice of the field F. Let the matrix $A = A_{(-1)}$. Then, $\mathrm{Rg}_C(E, A) = 2$ by Lemma 1 and Theorem 1. But $\mathrm{Rg}_R(E, A) > 2$, since the spectrum of A is non-real.

3. Upper bound of complexity of multiplying polynomials over a finite field

In [4], the rank of a finite-dimensional algebra $\mathscr{A}$ over the field F is defined as the rank of its structural tensor in some basis over F. Notice that the rank of the structural tensor of an algebra is independent of its basis. The rank of the algebra $\mathscr{A}$ is the complexity of solving the following problem: given the coefficients of the expansion, in some basis of the algebra $\mathscr{A}$ of two of its elements, it is required to find the coefficients of the expansion in this basis of their product. Strassen gave the two following inequalities. If $\mathscr{A}$ and $\mathscr{B}$ are algebras over F, then

$$\mathrm{Rg}(\mathscr{A}) + \mathrm{Rg}(\mathscr{B}) \geqslant \mathrm{Rg}(\mathscr{A} \oplus \mathscr{B}), \tag{3.1}$$

$$\mathrm{Rg}(\mathscr{A})\,\mathrm{Rg}(\mathscr{B}) \geqslant \mathrm{Rg}(\mathscr{A} \otimes_F \mathscr{B}) \tag{3.2}$$

(from inequality (3.2), the estimate in [2] can be obtained). Following [4], we also define the rank of a group G over the field F (denote it by $\mathrm{Rg}_F(G)$), as the rank over F of its group algebra FG.

Let us now estimate the complexity of the convolution or multiplication of polynomials. The complexity of this problem over the field F is equal to (see Proposition 3) the rank over the field F of the family of bilinear forms

$$P_n = \left\{ z_k = \sum_{0 \leqslant i, k-i \leqslant n} x_i y_{k-i} : 0 \leqslant k \leqslant 2n \right\}$$

in non-commuting variables x_k, y_k, $0 \leqslant k \leqslant n$. The computational significance of this problem is: given the coefficients of two polynomials

$$\sum_{0 \leqslant k \leqslant n} x_k t^k \quad \text{and} \quad \sum_{0 \leqslant k \leqslant n} y_k t^k$$

compute with the aid of n.c. the coefficients of their product

$$\sum_{0 \leqslant k \leqslant 2n} z_k t^k.$$

Denote by Z_m the cyclical group of order m.

Theorem 3

Given any field F and any n,

$$\mathrm{Rg}_F(Z_n) \leqslant \mathrm{Rg}_F(P_{n-1}) \leqslant \mathrm{Rg}_F(Z_{2n-1}). \tag{3.3}$$

Proof. Let us prove the left-hand inequality. Let c be a generator of Z_n and let

$$a=\sum_{0\leqslant i<n}\alpha_i c^i, \qquad b=\sum_{0\leqslant i<n}\beta_i c^i$$

be elements of $F(Z_n)$. Let

$$\gamma_k=\sum_{0\leqslant i,k-i<n}\alpha_i\beta_{k-i}, \qquad 0\leqslant k<2n-1.$$

We have the identity

$$ab=\sum_{0\leqslant k<n-1}(\gamma_k+\gamma_{k+n})c^k+\gamma_{n-1}c^{n-1}.$$

Hence $\mathrm{Rg}_F(Z_n)=\mathrm{Rg}(\{\gamma_k+\gamma_{k+n}: 0\leqslant k<n-1\}, \gamma_{n-1})\leqslant\mathrm{Rg}_F(\gamma_0,\ldots,\gamma_{2n-2})=\mathrm{Rg}_F(P_{n-1})$.

Let us prove the right-hand inequality. Let

$$f=\sum_{0\leqslant i<n}\alpha_i t^i, \qquad g=\sum_{0\leqslant i<n}\beta_i t^i\in F[t].$$

Let c be a generator of Z_{2n-1}. We define a mapping of the set of polynomials of $F[t]$ of degree not exceeding $2n-2$ in accordance with the expression

$$h=\sum_{0\leqslant i<2n-1}\eta_i t^i\to a_h=\sum_{0\leqslant i<2n-1}\eta_i c^i.$$

Then,

$$a_{fg}=a_f a_g=\sum_{0\leqslant i<2n-1}\gamma_i c^i.$$

Hence $\mathrm{Rg}_F(P_{n-1})=\mathrm{Rg}_F(\gamma_0,\ldots,\gamma_{2n-2})\leqslant\mathrm{Rg}_F(Z_{2n-1})$. The lemma is proved.

We are interested in the complexity P_n over a finite field, whose characteristic we shall henceforth denote in this section by q. We denote the finite field containing q^r elements by $F(q^r)$.

Lemma 4

For every n,

$$\mathrm{Rg}_{F(q)}(F(q^n))\leqslant\mathrm{Rg}_{F(q)}(P_{n-1})\leqslant\mathrm{Rg}_{F(q)}(F(q^{2n-1})).$$

Proof. Let χ be the primitive element of the field $F(q^n)$ over $F(q)$. For any $k\geqslant n$, $\theta_{k,0},\ldots,\theta_{k,n-1}\in F(q)$, exist such that

$$\chi^k=\sum_{0\leqslant i<n}\theta_{ki}\chi^i.$$

Let

$$\zeta=\sum_{0\leqslant i<n}\alpha_i\chi^i, \qquad \nu=\sum_{0\leqslant i<n}\beta_i\chi^i, \qquad \alpha_i,\beta_i\in F(q).$$

Then

$$\gamma_k=\sum_{0\leqslant i,k-i<n}\alpha_i\beta_{k-i}, \qquad 0\leqslant k<2n-1.$$

Then

$$\mathrm{Rg}_{F(q)}(F(q^n))=\mathrm{Rg}_{F(q)}\left(\left\{\gamma_i+\sum_{n\leqslant k<2n-1}\theta_{ki}\gamma_k : 0\leqslant i<n\right\}\right)$$
$$\leqslant\mathrm{Rg}_{F(q)}(\gamma_0,\ldots,\gamma_{2n-2})=\mathrm{Rg}_{F(q)}(P_{n-1}).$$

The right-hand inequality of (3.4) is proved in the same way as the right-hand inequality (3.3). Inequalities (3.3) and (3.4) are proved in a general form in [6]. The next lemma follows from the residue theorem.

Lemma 5

Let n and q be relatively prime, and let

$$t^n-1=\prod_i f_i,$$

where the polynomials f_i are irreducible over $F(q)$ and pairwise distinct. Then,

$$F(q)(Z_n)\simeq\sum_i\oplus F(q^{k_i}), \qquad k_i=\deg f_i.$$

The next lemma is familiar in the theory of finite fields.

Lemma 6

Let $r=q^m$. Then,

$$t^r-t=\prod_i f_i,$$

where the product is taken (once each) over all polynomials f_i, irreducible over $F(q)$, whose degree is a devisor of m.

We define the function $g_q(n)$ as follows. We put $g_q(2)=g_q(3)=2$, $g_q(4)=3$. If $m=[(q^{s-1})/2]+1>4$ for some integer s, we put $g_q(m)=2qg_q(s)$. If the integer $m>4$ is such that $[(q^s)/2]<m\leqslant[(q^{s+1})/2]$ for some integer s, we put $g_q(m)=g_q([(q^s)/2]+1)$. The function $g_q(n)$ is correctly defined and is non-decreasing with respect to n.

Theorem 4

For every n and prime q,

$$\mathrm{Rg}_{F(q)}(P_{n-1})\leqslant ng_q(n).$$

Proof. Inequality (3.5) can be checked directly for $n = 2, 3, 4$. We then use induction on n. Let (3.5) hold for $n\leqslant s$. We put $r=q^s-1$, and by Lemmas 5 and 6,

$$F(q)(Z_r)\simeq\sum_i \oplus F(q^{k_i}),$$

where

$$\sum_i k_i=r$$

and every k_i is a divisor of s. Hence, by inequality (3.1), we obtain the inequality

$$\mathrm{Rg}_{F(q)}(Z_r)\leqslant\sum_i \mathrm{Rg}_{F(q)}(F(q^{k_i})).$$

Using inequalities (3.3), (3.6), (3.4) successively, along with our inductive assumption, the monotonicity of $g_q(n)$ with respect to n, and Lemma 6, we obtain the chain of inequalities

$$\mathrm{Rg}_{F(q)}(P_{[(r-1)/2]})\leqslant\mathrm{Rg}_{F(q)}(Z_r)\leqslant\sum_i \mathrm{Rg}_{F(q)}(F(q^{k_i}))$$

$$\leqslant\sum_i \mathrm{Rg}_{F(q)}(P_{k_i-1})\leqslant\sum_i k_ig_q(k_i)\leqslant g_q(s)\sum_i k_i=rg_q(s).$$

Let the integer $n>4$ be such that $[q^{s-1}/2]<n\leqslant[q^s/2]$. Let us prove (3.5) for this n. The rank $\mathrm{Rg}_{F(q)}(P_n)$ is not monotonically decreasing with respect to n. Using this fact, inequality (3.7), and the definition of the function $g_q(n)$, we obtain the chain of inequalities

$$Rq_{F(q)}(P_{n-1})\leqslant\mathrm{Rg}_{F(q)}(P_{[(r-1)/2]})\leqslant r\,g_q(s)$$
$$\leqslant([q^{s-1}/2]+1)\,g_q([q^{s-1}/2]+1)\leqslant ng_q(n).$$

The theorem is proved.

4. A group of linear tensor-rank-preserving transformations

Let us describe a group of non-singular linear transformations of the tensor product of vector spaces, that preserve the rank. Obviously, rank is preserved under the following two types of transformations of tensor product $V_1\otimes\ldots\otimes V_k$ (we assume that $\dim V_i>1$, $1\leqslant i\leqslant k$, otherwise, the one-dimensional factor can be discarded without loss of generality).

(a) $T_1\otimes\ldots\otimes T_k$, where T_i is a non-singular linear transformation of V_i, $1\leqslant i\leqslant k$;

(b) if $T: V_i\to V_j$ is, for certain i, j, an isomorphism of linear spaces, then rank is preserved under transformations of the type

$$v_1\otimes\ldots\otimes v_i\otimes\ldots\otimes v_j\otimes\ldots\otimes v_k\to v_1\otimes$$
$$\ldots\otimes T^{-1}(v_j)\otimes\ldots\otimes T(v_i)\otimes\ldots\otimes v_k.$$

A transformation of type (b) will be called a transposition, by analogy with ordinary matrices.

Theorem 5

The group of all non-singular linear transformations of space $V_1\otimes\ldots\otimes V_k$, preserving rank, is the same as the group generated by transformations of types (a), (b).

We require a preliminary lemma.

Lemma 7

Let $\tau_1, \tau_2\in V_1\otimes\ldots\otimes V_k$ and for every $\alpha,\beta\in F$, let $\mathrm{Rg}(\alpha\tau_1+\beta\tau_2)\leqslant 1$. Then, for some i, $1\leqslant i\leqslant k$, vectors $v^{(1)}, v^{(2)}\in V_i$ u $v_j\in V_j$, $j\neq i$, exist, such that, for $m = 1, 2$, we have the equations

$$\tau_m=v_1\otimes\ldots\otimes v_{i-1}\otimes v^{(m)}\otimes v_{i+1}\otimes\ldots\otimes v_k.$$

Proof. The tensors τ_1, τ_2 can be written as

$$\tau_1=u_1\otimes\ldots\otimes u_k,\qquad \tau_2=w_1\otimes\ldots\otimes w_k. \tag{4.1}$$

Assume that, among the pairs u_i, w_i, $1\leqslant i\leqslant k$, there are two, say with $i = 1, 2$, in which the vectors are linearly independent; otherwise, (4.1) are the required equations. Let $\tau_1+\tau_2=v_1'\otimes v_2'\ldots$. An element $f\in V^*_1$ exists such that $f(v'_1)=0$, but $f(u_1)\neq 0$ and $f(w_1)\neq 0$. From (4.1) we then obtain $f(u_1)u_2\otimes\ldots\otimes u_k+f(w_1)w_2\otimes\ldots\otimes w_k=0$. Hence u_2 and w_2 are linearly dependent. This contradiction proves the lemma.

We turn to the proof of Theorem 5. We shall confine ourselves to the case $k = 3$. Let $\{u_i\}$, $\{v_j\}$, $\{w_s\}$ be bases of U, V, W respectively, and let $\dim W\geqslant\dim V\geqslant\dim U\geqslant 2$. Let $\mathcal{B}_q=\{u_i\otimes v_j\otimes w: w\in W\}$. Let T be the considered linear transformation of space $U\otimes V\otimes W$, preserving rank. We apply Lemma 7 to the tensors $T(u_1\otimes v_1\otimes w_1)$ and $T(u_1\otimes v_1\otimes w_2)$ and find that, for example the following relations hold:

$$T(u_1\otimes v_1\otimes w_1)=u^{11}\otimes v^1\otimes w^{11},\qquad T(u_1\otimes v_1\otimes w_2)=u^{11}\otimes v^2\otimes w^{11}.$$

Then, for every $w \in W$, $v \in V$ exists such that $T(u_1 \otimes v_1 \otimes w) = u^{11} \otimes \otimes v \otimes w^{11}$. For the proof, we apply Lemma 7 to the pair of tensors $T(u_1 \otimes v_1 \otimes w_m)$ and $T(u_1 \otimes v_1 \otimes w)$ with $m = 1, 2$ (recalling that T is injective, and v^1 linearly independent of v^2). In this case, $\dim T(\mathcal{P}_{11}) = \dim W \geqslant \dim V$, and hence $\dim V = \dim W$; then, possibly after transposing V and W, we can assume that

$$T(\mathcal{P}_{11}) = \{u^{11} \otimes v^{11} \otimes w:\ w \in W\}.$$

We consider $\mathcal{P}_{1l}$, $l \neq 1$, and using similar arguments to those for $\mathcal{P}_{11}$, we assume, for example that

$$T(\mathcal{P}_{11}) \subseteq \{u^{11} \otimes v \otimes w^{11}:\ v \in V\}.$$

We will show that this is impossible, and hence

$$T(\mathcal{P}_{11}) = \{u^{11} \otimes v^{11} \otimes w:\ w \in W\}.$$

We apply Lemma 7 to the pair of tensors $T(u_1 \otimes v_1 \otimes w_1)$, $T(u_1 \otimes v_1 \otimes w_1)$; $T(u_1 \otimes v_1 \otimes w_2)$; $T(u_1 \otimes v_2 \otimes (w_1 + w_2))$, $T(u_1 \otimes v_1 \otimes (w_1 + w_2))$. There exist $w^{(1)}$, $w^{(2)} \in W$, $v^{(1)}$, $v^{(2)} \in V$, such that, by (4.2) and (4.3):

$$T(u_1 \otimes v_1 \otimes w_1) = u^{11} \otimes v^{11} \otimes w^{(1)}, \qquad T(u_1 \otimes v_1 \otimes w_1) = u^{11} \otimes v^{(1)} \otimes w^{11},$$

$$T(u_1 \otimes v_1 \otimes w_2) = u^{11} \otimes v^{11} \otimes w^{(2)}, \qquad T(u_1 \otimes v_1 \otimes w_2) = u^{11} \otimes v^{(2)} \otimes w^{11}.$$

Below, in the proof of the theorem, the relation of non-zero vectors $v \approx u$ means that v and u are linearly dependent. By Lemma 7 and Eqs. (4.5), either $w^{(1)} \approx w^{11}$ and $v^{11} \approx v^{11}$, or $w^{(2)} \approx w^{11}$ and $v^{11} \approx v^{(1)}$, but not simultaneously. In either case we obtain a contradiction, justifying our assumption (4.3), and proving inclusion (4.4).

Now, on varying i with fixed l, and using similar arguments, we obtain, for all pairs i, l, the inclusion

$$T(\mathcal{P}_{il}) \subseteq \{u^{il} \otimes v^{il} \otimes w;\ w \in W\}.$$

It is also clear from Eq. (4.4) (using Lemma 7) that either $u^{1j} \approx u^{11}$, or $v^{1j} \approx v^{11}$, but not both simultaneously ($j \neq l$). If we assume that $v^{1j} \approx v^{11}$, then, by varying j, we find (using inclusion (4.6)) that, for any j, this condition is satisfied. But then, by Lemma 7 and inclusion (4.6), we have $T(\{u_1 \otimes v \otimes w_1:\ v \in V\}) \subseteq \{u \otimes v^{11} \otimes w^{(1)}:\ u \in U\}$, but $\dim T(\{u_1 \otimes v \otimes w_1:\ v \in V\}) = \dim V \geqslant \dim U$, and hence in this case $\dim V = \dim U$. Using transposition of U and V, we can assume that $u^{1j} \approx u^{11} \approx u^1$. Let us show that, for any i, j, l, we have $u^{ij} \approx u^{il}$. Assume that this is not the case for certain i, j, l ($i \neq 1$, $j \neq l$). We then have the equations

$$T(u_l \otimes v_j \otimes w_1) \approx u^{ij} \otimes v^{il} \otimes w^{(1)}, \qquad T(u_l \otimes v_l \otimes w_1) \approx u^{il} \otimes v^{il} \otimes w^{(1)},$$

$$T(u_1 \otimes v_j \otimes w_1) \approx u^{1} \otimes v^{ij} \otimes w^{(1)}, \qquad T(u_1 \otimes v_l \otimes w_1) \approx u^{1} \otimes v^{il} \otimes w^{(1)}.$$

Then, by Lemma 7, $v^{ij} \approx v^{il}$. Also, from (4.7) and Lemma 7, we find that either

$$u^{ij} \approx u^{1}, \qquad v^{il} \approx v^{il}, \tag{4.8}$$

or else $u^{il} \approx u^{1}$ and $v^{ij} \approx v^{il}$. Assume that, say, relations (4.8) hold. Applying Lemma 7 to the tensor pair $T(u_l \otimes (v_j + v_l) \otimes w_1)$ and $T(u_1 \otimes (v_j + v_l) \otimes w_1)$ and using (4.8), we obtain a contradiction. This shows that, for all i, the vector $u^{ij} \approx u^{i}$ is independent of j, while $v^{ij} \approx v^{il}$ for all $i \neq j$. It then follows from (4.7) that $v^{ij} \approx u^{il} \approx v^{j}$ for all i, j. All in all, we find that

$$T(\mathcal{P}_{ij}) \subseteq \{u^{i} \otimes v^{j} \otimes w:\ w \in W\}. \tag{4.9}$$

Now, $i \neq l$. From inclusions (4.9) we have

$$T(u_i \otimes v_j \otimes w_p) = u^{i} \otimes v^{j} \otimes w', \qquad T(u_l \otimes v_l \otimes w_p) = u^{i} \otimes v^{i} \otimes w''.$$

By Lemma 7, $w' \approx w''$, and hence, for any i, j (varying j, then varying i with fixed j), we have $T(u_i \otimes v_j \otimes w_p) \approx u^{i} \otimes v^{j} \otimes w^{p}$ for some w^p.

In short, we have shown that $T(u_i \otimes v_j \otimes w_p) = \varkappa_{ijp} u^{i} \otimes v^{j} \otimes w^{p}$, $\varkappa_{ijp} \in F$, where $\{u^i\}$, $\{v^j\}$, $\{w^p\}$ are bases of spaces U, V, W respectively. The transformation T is linear and preserves rank, and hence

$$1 = \operatorname{Rg}\left(\sum_{i,j,p} u_i \otimes v_j \otimes w_p\right) = \operatorname{Rg}\left(\sum_{i,j,p} \varkappa_{ijp} u^{i} \otimes v^{j} \otimes w^{p}\right).$$

Hence it follows that non-zero $\alpha_i, \beta_j, \gamma_p \in F$, exist, such that

$$\left(\sum_i \alpha_i u^{i}\right) \otimes \left(\sum_j \beta_j v^{j}\right) \otimes \left(\sum_p \gamma_p w^{p}\right) = \sum_{i,j,p} \varkappa_{ijp} u^{i} \otimes v^{j} \otimes w^{p}.$$

Starting from this, we find that the transformation T is the same, up to possible transformations arising during its changes during the proof of the theorem, as the tensor product of non-singular transformations T_1, T_2, T_3 in U, V, W respectively, such that $T_1(u_i) = \alpha_i u^{i}$, $T_2(v_j) = \beta_j v^{j}$, $T_3(w_p) = \gamma_p w^{p}$ with respect to all i, j, p. This proves the theorem.

Notice that, in the proof of the theorem, it was only demanded of the transformation T that it be linear and one-to-one, and that it map tensors of rank 1 into tensors of rank 1. For matrices ($k = 2$), a result similar to Theorem 5 was obtained in [16].

5. Tensors of large rank

We shall examine the behaviour of rank almost everywhere, and prove that tensors of large rank exist with coefficients from the set $\{0, 1\}$.

Below, q denotes a prime or zero. Let the family $X^{(q)}=\{X^{(q)}(i_1,\ldots,i_k);\ 1\leqslant i_j\leqslant n_j,\ 1\leqslant j\leqslant k\}$ be algebraically independent over $\overline{F(q)}$ (here and henceforth, $\bar{F}$ is the algebraic closure of the field F). Denote by $H^{(q)}=\overline{F(q)(X^{(q)})}$ the field of algebraic functions. We define $r_q(n_1,\ldots,n_k)=\mathrm{Rg}_{H^{(q)}}(\mathrm{T}^{(q)}(n_1,\ldots,n_k))$, where the tensor $\mathrm{T}^{(q)}(n_1,\ldots,n_k)\in(H^{(q)})^{n_1}\otimes_{H^{(q)}}\ldots\otimes_{H^{(q)}}(H^{(q)})^{n_k}$, and $(i_1,\ldots,i_k)$ is an element of the tensor $\mathrm{T}^{(q)}(n_1,\ldots,n_k)$ and equal to $X^{(q)}(i_1,\ldots,i_k)$ (for uniformity, we assume that $F(0)=\mathcal{Q}$).

Lemma 8

For the algebraically closed field F, the rank (over F) of tensors of $F^{n_1}\otimes\ldots\otimes F^{n_k}$ is almost everywhere equal to $r_q(n_1,\ldots,n_k)$.

For the proof, we shall confine ourselves to the case $k=3$ and consider the following system of algebraic equations in the variables $x^{(l)}(i_1,\ 1)$, $x^{(l)}(i_2,\ 2)$, $x^{(l)}(i_3,\ 3)$ ($\tau(i_1,i_2,i_3)$ are parameters):

$$\sum_{1\leqslant l\leqslant N} x^{(l)}(i_1,1)x^{(l)}(i_2,2)x^{(l)}(i_3,3)=\tau(i_1,i_2,i_3),$$

$$1\leqslant i_1\leqslant n_1,\quad 1\leqslant i_2\leqslant n_2,\quad 1\leqslant i_3\leqslant n_3$$

(N is fixed for the present). The solvability of this system is equivalent to the fact that the rank of the tensor $\tau\in F^{n_1}\otimes F^{n_2}\otimes F^{n_3}$, composed from $\{\tau(i_1,\ i_2,\ i_3)\}$, is not greater than N.

Let us apply to system (5.1) the procedure (see [17], Chapter 11) for eliminating variables. As a result, non-zero polynomials $f_{il}(\{\tau(i_1,\ i_2,\ i_3)\})$ with coefficients from the field $F(q)$ will exist, such that the solvability of system (5.1) is equivalent to the following disjunction of conjunctions:

$$\bigvee_i (f_{i1}=0\wedge\ldots\wedge f_{ip}=0\wedge f_{i,p+1}\neq 0\wedge\ldots\wedge f_{is}\neq 0).$$

We take the least N such that $p=0$ in this expression for some i. This N is in fact equal to $r_q(n_1,\ n_2,\ n_3)$, since we can take as $\tau(i_1,\ i_2,\ i_3)$ algebraically independent $X^{(q)}(i_1,\ i_2,\ i_3)$. On the other hand, for the tensor $\tau=(\tau(i_1,\ i_2,\ i_3))$, whose elements satisfy the inequalities $f_{i1}(\{\tau(i_1,\ i_2,\ i_3)\})\neq 0,\ \ldots,\ f_{is}(\{\tau(i_1,\ i_2,\ i_3)\})\neq 0$, we have $\mathrm{Rg}_F(\tau)\leqslant N$, while the set of the $\sigma\in F^{n_1}\otimes F^{n_2}\otimes F^{n_3}$, for which $\mathrm{Rg}_F(\sigma)<N$, lies in a proper closed subset of space $F^{n_1}\otimes F^{n_2}\otimes F^{n_3}$ by virtue of the choice of N (we recall that a Zariski topology is considered). The lemma is proved.

It follows from Theorem 2 that $r_q(2,\ m,\ n)=\min\{n,\ 2m\}$, where $m\leqslant n$. Let us apply the lower bound to $r_q(n_1,\ldots n_k)$.

Proposition 4 $r_q(n_1,\ldots,\ n_k)\geqslant(n_1\ldots n_k)/(n_1+\ldots+n_k-k+1)$.

We put $r=r_q(n_1,\ldots,\ n_k)$ and consider the mapping R: $F^{r(n_1+\ldots+n_k)}\to F^{n_1}\otimes\ldots\otimes F^{n_k}$, setting

$$R(\{x^{(l)}(i_j,j):\ 1\leqslant l\leqslant r,\ 1\leqslant i_j\leqslant n_j,\ 1\leqslant j\leqslant k\})$$
$$=\left(\left\{\sum_{1\leqslant l\leqslant r} x^{(l)}(i_1,1)\ldots x^{(l)}(i_k,k):\ 1\leqslant i_j\leqslant n_j,\ 1\leqslant j\leqslant k\right\}\right).$$

In the space $F^{r(n_1+\ldots+n_k)}$ we consider the family of sets $\mathcal{L}=\{S\}$, specified by the inequalities

$$\begin{aligned} &x^{(1)}(i_{11},\ 1)\ldots x^{(1)}(i_{1,k-1},\ k-1)\neq 0,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ &x^{(r)}(i_{r1},\ 1)\ldots x^{(r)}(i_{r,k-1},\ k-1)\neq 0. \end{aligned}\tag{5.2}$$

Each S depends on the indices $i_{11},\ldots,i_{1,k-1},\ i_{21},\ldots,i_{r,k-1}$, $1\leqslant i_{lj}\leqslant n_j$. If

$$\tau\notin\bigcup_{S\in\mathcal{L}} S,$$

then $\mathrm{Rg}_F(R(\tau))<r$, and hence

$$\overline{R(\bigcup S)}=F^{n_1}\otimes\ldots\otimes F^{n_k}$$

By Lemma 8. From this, using Chevalley's theorem, we find that $S\in\mathcal{L}$ exists, specified by inequalities (5.2) (assume that, for this S, all $i_{lj}=1$), such that $R(S)$ contains a non-empty open set (we say that the mapping R: $S\to F^{n_1}\otimes\ldots\otimes F^{n_k}$ is "almost into").

We construct a mapping φ: $S\to F^{r(n_1+\ldots+n_k-k+1)}$ according to the following rule. For convenience, we divide the variables $x^{(l)}(i_j,\ j)$ into blocks with respect to the index l, $1\leqslant l\leqslant r$, and we define the mapping φ_l in the l-th block:

$$\varphi_l(x^{(l)}(1,1),\ldots,x^{(l)}(n_1,1);\ldots;x^{(l)}(1,k-1),$$
$$\ldots,x^{(l)}(n_{k-1},\ k-1);\ x^{(l)}(1,\ k),\ldots,x^{(l)}(n_k,\ k))$$
$$=(x^{(l)}(2,\ 1)/x^{(l)}(1,\ 1),\ldots,x^{(l)}(n_1,\ 1)/x^{(l)}(1,\ 1);$$
$$\ldots;x^{(l)}(2,\ k-1)/x^{(l)}(1,\ k-1),\ldots,x^{(l)}(n_{k-1},\ k-1)/x^{(l)}(1,\ k-1);\ x^{(l)}(1,\ k)x^{(l)}(1,\ 1)$$
$$\ldots x^{(l)}(1,\ k-1),\ldots,x^{(l)}(n_k,\ k)x^{(l)}(1,\ 1)\ldots x^{(l)}(1,\ k-1)).$$

By collecting all r blocks of variables, we obtain the mapping φ.

Let I: $F^{r(n_1+\ldots+n_{k-1}+1)}\to F^{r(n_1+\ldots+n_k)}$ be the imbedding (we define it in the l-th block of variables $x^{(l)}(i_j,\ j)$):

$$I_t(x^{(t)}(2,\ 1),\ldots,x^{(t)}(n_1,\ 1);\ldots;x^{(t)}(2,\ k-1),$$
$$\ldots,x^{(t)}(n_{k-1},\ k-1);\ x^{(t)}(1,\ k),\ x^{(t)}(2,\ k),\ldots$$
$$\ldots,x^{(t)}(n_k,\ k))=(1,\ x^{(t)}(2,\ 1),\ldots,x^{(t)}(n_1,\ 1);$$
$$\ldots;1,\ x^{(t)}(2,\ k-1),\ldots,x^{(t)}(n_{k-1},\ k-$$
$$-1);\ x^{(t)}(1,\ k)x^{(t)}(2,\ k),\ldots,x^{(t)}(n_k,\ k)).$$

Then, the following diagram is commutative:

$$\begin{array}{ccc} S & \xrightarrow{\varphi} & F^{r(n_1+\ldots+n_k-k+1)} \\ R\downarrow & & \downarrow I \\ F^{n_1}\otimes\ldots\oplus F^{n_k} & \xleftarrow{R} & F^{r(n_1+\ldots+n_k)} \end{array}$$

The left-hand arrow is a mapping "almost into" and hence the composition of the right-hand and lower arrows is also a mapping "almost into", and hence $r\geqslant(n_1\ldots n_k)/(n_1+\ldots+n_k-k+1)$. The proposition is proved.

Using the method described in [18], we can easily construct tensors of rank close to the lower bound of Proposition 4. In Theorem 6 (below) we prove the (ineffective) existence of tensors of $F(q)^{n_1}\otimes\ldots\otimes F(q)^{n_k}$, with coefficients from the set $\{0,\ 1\}$ and with rank differing from the maximum possible (for the given $n_1,\ldots,n_k$) by not more than a factor dependent only on k. The method of proving the theorem is similar to the method used in [15], for constructing "hard-to-compute" polynomials with coefficients from the set $\{0,\ 1\}$. However, if we employ directly the idea used in [19], we obtain a lower bound of order $r_q(n_1,\ldots,n_k)/\ln r_q(n_1,\ldots,n_k)$ (for polynomials of degree n the method described in [15] gives a greatest lower bound of order $n/\ln n$); therefore, for the proof of Theorem 6, we have to bring in extra considerations. Also, as distinct from [19], the characteristic q is in our case arbitrary.

Theorem 6

For any $n_1,\ldots,n_k$ there exists a tensor $\tau\in F(q)^{n_1}\otimes\ldots\otimes F(q)^{n_k}$ with coefficients from the set $\{0,\ 1\}$, such that

$$\mathrm{Rg}_{F(q)}\tau>\frac{n_1\ldots n_k}{5(n_1+\ldots+n_k)\log_2 k}.$$

Proof. In the proof we shall confine ourselves to the case $k=3$, which is the most important for us (it corresponds, as mentioned in the introduction, to the case of a family of bilinear forms). We put $r=[n_1n_2n_3/7(n_1+n_2+n_3)]$ and assume that every tensor of space $F(q)^{n_1}\otimes F(q)^{n_2}\otimes F(q)^{n_3}$ with coefficients from the set $\{0,\ 1\}$ has rank not greater than r.

We consider the system of equations (5.1) with $N=r$, and denote by $q_1,\ldots,q_{n_1n_2n_3}$ the left-hand sides of the equations (numbered in any order).

Lemma 9

Let the degree of each of polynomials $h_1,\ldots,h_m\in F(q)[y_1,\ldots,y_n]$ be not greater than d, and let $2^m>\binom{n+dm}{n}$. Then, there exists a non-zero polynomial $H\in F(q)[z_1,\ldots,z_m]$ of degree not greater than 1 with respect to each of the variables z_i, $1\leqslant i\leqslant m$, such that $H(h_1,\ldots,h_m)\equiv 0$.

For the proof, we consider the 2^m coefficients of the polynomial $H(z_1,\ldots,z_m)$, with all possible monomials of the type $z_{i_1}\ldots z_{i_l}$, $i_1<\ldots<i_l$, as unknowns. We equate to zero all the coefficients of monomials of variables $y_1,\ldots,y_n$ in the polynomial $H(h_1(y_1,\ldots,y_n),\ldots,h_m(y_1,\ldots,y_n))$; as a result, we obtain a system of not more than $\binom{n+dm}{n}$ equations (over $F(q)$) homogeneous equations in 2^m unknowns. The non-trivial solution of this system gives the coefficients of polynomial H. The lemma is proved.

Let us show that Lemma 9 is applicable to polynomials $g_1,\ldots,g_{n_1n_2n_3}$ for $r=r(n_1+n_2+n_3)$, $m=n_1n_2n_3$. For this, it is sufficient to verify the inequality $2^m>\binom{n+dm}{n}$. We have $\binom{n+3m}{n}<\binom{4m}{n}$, since $n<m$. Further, by Stirling's formula, $\binom{4m}{n}<\left(\frac{12m}{n}\right)^n$. The inequality $12m/n<2^{m/n}$ holds for $m/n\geqslant 7$, while the latter holds by the definition of r.

On applying the lemma, we obtain a polynomial $H(z_1,\ldots,z_{n_1n_2n_3})$. By Lemma 9 and the above assumption, given any substitution into H of zero or unity for each of the variables z_i, $1\leqslant i\leqslant m$, we obtain a zero value. We write the polynomial H as $z_mu_m+v_m$, where $u_m,v_m\in F(q)[z_1,\ldots,z_{m-1}]$ and are at most of first degree with respect to each of the variables z_i, $1\leqslant i\leqslant m-1$. Each of the polynomials u_m, v_m vanishes for any substitution of zero or unity for each of the variables z_i, $1\leqslant i\leqslant m-1$. One of polynomials u_m, v_m does not vanish identically, and hence it has the same properties as the polynomial H, but contains one variable less. We apply to it the same procedure as above, and eliminate the variable z_{m-1}, and so on. At the end of the process, we arrive at a non-zero polynomial of a single variable z_1 of at most the first degree, having two roots, zero and unity. This contradiction proves the theorem.

The author thanks A. O. Slisenko for unfailing interest, and also A. M. Vershik for valuable observations.

Translated by D. E. Brown.

REFERENCES

1. BORODIN, R., and MUNRO, I., *The computational complexity of algebraic and numeric problems*, Amer. Elsevier, New York, 1975.
2. STRASSEN, V. Gauss's algorithm is not optimal, in: *Cybernetic Collection* (Kibernetich. sb.), New series, No. 7, 67–70, Mir, Moscow, 1970.
3. WINOGRAD, S. On the number of multiplications necessary to compute certain functions, *Comm. Pure Appl. Math.*, 23, 165–179, 1970.
4. STRASSEN, V., Vermeidung von Divisionen, *J. Reine angew. Math.*, 264, 184–202, 1973.
5. GRIGOR'EV, D. Yu., On the algebraic complexity of computing a pair of bilinear forms, *Zap. nauch. seminarov LOMI Akad. Nauk SSSR*, 47, 159–163, 1974.
6. FIDUCCIA, C. M., and ZALCSTEIN, Y., Algebras having linear multiplicative complexity. *J. Assoc. Comput. Machinery*, 24, No. 2, 311–331, 1977.
7. KNUT, D. E., *The art of programming for computers*, vol. 2, Mir, Moscow, 1977.
8. KARATSURA, A. A., and OFMAN, Yu. P., Multiplication of *n*-ary numbers in automata, *Dokl. Akad. Nauk SSSR*, 145, no. 2, 293–294, 1962.
9. SHENKHAGE, A., and STRASSEN, V., Fast multiplication of large numbers, in: *Cybernetics collection* (new series), No. 10, 87–98, 1973.
10. SCHÖNHAGE, A., Schnelle Multiplikation von Polynomen über Körpern der Charakteristik, 2, *Acta Inform.*, 7, No. 4, 395–398, 1977.
11. YABLONSKII, S. V., Basic concepts of cybernetics, *Probl. kibernetiki*, No. 2, 7–38, 1959.
12. BOREL, A., *Linear algebraic groups* (Russian translation), Mir, Moscow, 1972.
13. GANTMAKHER, F. R., *Theory of matrices* (Teoriya matrits), Gostekhizdat, Moscow, 1954.
14. GRIGOR'EV, D. Yu., Some new bounds on tensor rank, LOMI Preprint E-2-78, 1978.
15. JA' JA', J., Optimal evaluation of pairs of bilinear forms, *Proc. 10th Ann ACM Symp. Th. Comput., San Diego Calif.*, 173–183, 1978.
16. MARCUS, M., and MOYLS, B., Transformations on tensor product spaces, *Pcif. J. Math.*, 9, No. 4, 1215–1221, 1959.
17. VAN DER WAERDEN, B., *Modern algebra* (Russian translation), ONTI, Moscow-Leningrad, 1937.
18. STRASSEN, V., Polynomials with rational coefficients which are hard to compute, *SIAM J. Comput.*, 3, No. 2, 128–149, 1974.
19. LIPTON, R., Polynomials with 0–1 coefficients that are hard to evaluate, *SIAM J. Comput.*, 7, No. 1, 61–69, 1978.

U.S.S.R. Comput. Maths. Math. Phys. Vol. 19, pp. 21–33 0041–5553/79/0601–0021$07.50/0
Pergamon Press Ltd. 1980. Printed in Great Britain.

EXTENSION OF STRASSEN'S ESTIMATE TO THE SOLUTION OF ARBITRARY SYSTEMS OF LINEAR EQUATIONS*

V. I. SOLODOVNIKOV

Moscow

(*Received* 13 *February* 1978; *revised* 6 *July* 1978)

It is shown that any system of linear equations can be solved in $O(\max(m, n)(\min(m, n))^{1.81})$ operations, where m is the number of equations and n the number of unknowns.

Our main result below is to construct an algorithm for solving any system of m linear equations with n unknowns in $O(\max(m, n)(\min(m, n))^{\log 3.5})$ operations (all the logarithms are to base 2, so that $\log 3.5 \approx 1.81$; by operation we mean what is called a step in [1], though all the estimates still hold if an operation is taken to mean simply an arithmetic operation). Gauss's algorithm for solving this problem requires $O(\max(m, n) \times (\min(m, n))^2)$ operations.

An algorithm is constructed in [2] for computing the product of any two $n \times n$ matrices in $O(n^{\log 7})$ operations ($\log 7 \approx 2.81$), and has come to be known as fast matrix multiplication (the ordinary algorithm requires $O(n^3)$ operations). Also in [2] an algorithm is given for solving a non-degenerate system of n linear equations with n unknowns, again in $O(n^{\log 7})$ operations. It was shown in [3], however, that this algorithm is not applicable to every square non-degenerate system.

The so-called LUP expansion of a matrix is applied in [1, 3] for solving a system of linear equations. Using fast matrix multiplication, an algorithm is constructed for computing the LUP expansion of an $n \times n$ matrix in $O(n^{\log 7})$ operations. Although this algorithm is not applicable to every matrix, it is applicable to any non-degenerate matrix. It is shown by means of the algorithms that the solution of any system of n linear equations with n unknowns, such that the matrix of coefficients of the system is non-degenerate, is possible in $O(n^{\log 7})$ operations.

1. LUP expansion of matrices

All the matrices considered below are matrices over the same arbitrary field K.

Definition 1. The LUP expansion of the $m \times n$ matrix A, $m \leqslant n$, is defined as the triple of matrices (L, U, P) such that L is a lower-triangular $m \times m$ matrix, any element of the principal diagonal of which is either zero or unity, U is an upper triangular $m \times n$ matrix with non-zero elements on its principal diagonal, and P is a permutable $n \times n$ matrix, while $A = L.\ U.\ P.$

Zh. vychisl. Mat. mat. Fiz., 19, 3, 581–593, 1979.