A $\frac{5}{2}n^2$–Lower Bound for the Rank of $n \times n$–Matrix Multiplication over Arbitrary Fields

Markus Bläser
Institut für Informatik II, Universität Bonn
Römerstr. 164, 53117 Bonn, Germany
blaeser@cs.uni-bonn.de

Abstract

We prove a lower bound of $\frac{5}{2}n^2 - 3n$ for the rank of $n \times n$–matrix multiplication over an arbitrary field. Similar bounds hold for the rank of the multiplication in noncommutative division algebras and for the multiplication of upper triangular matrices.

1. Introduction

One of the most studied quantities in algebraic complexity theory is the rank of $n \times n$–matrix multiplication. Roughly speaking, our problem is the following: Given $n \times n$–matrices $X = (X_{i,j})$ and $Y = (Y_{i,j})$ with indeterminates $X_{i,j}$ and $Y_{i,j}$ over some ground field k, what is the smallest number r of products $p_\rho = u_\rho(X)v_\rho(Y)$ with linear forms $u_\rho(X)$ in the indeterminates $X_{i,j}$ and linear forms $v_\rho(Y)$ in the indeterminates $Y_{i,j}$, such that the entries of XY are contained in the linear span of the p_ρ? This quantity r is called the bilinear complexity or the rank of $n \times n$–matrix multiplication.

A related quantity is the multiplicative complexity of $n \times n$–matrix multiplication. Here, we consider products of the form $p_\lambda = u_\lambda(X,Y)v_\lambda(X,Y)$, where $u_\lambda(X,Y)$ and $v_\lambda(X,Y)$ are linear forms in the $X_{i,j}$ and $Y_{i,j}$. The smallest number l of such products p_λ such that the entries of XY are contained in the linear span of the p_λ is called the multiplicative complexity of $n \times n$–matrix multiplication. Obviously, the multiplicative complexity is a lower bound for the rank, and it is not hard to see that twice the multiplicative complexity is an upper bound for the rank (see e.g. [6, Eq. 14.8]). In this paper we are solely concerned with the rank. For further background see [9, 6].

It is sometimes more convenient to use a coordinate-free definition of rank, see e.g. [6, Ch. 14.1]. In the following, if V is a vector space, let V^* denote the dual space of V, i.e., the vector space of all linear forms on V.

Definition 1 *Let k be a field, U, V, W finite dimensional vector spaces over k, and $\phi : U \times V \to W$ a bilinear map.*

1. *A bilinear computation of length r for ϕ is a sequence $\beta = (f_1, g_1, w_1, \dots, f_r, g_r, w_r)$, such that $f_\rho \in U^*$, $g_\rho \in V^*$, $w_\rho \in W$, and*
$$\phi(u,v) = \sum_{\rho=1}^{r} f_\rho(u)g_\rho(v)w_\rho \quad \text{for all } u \in U, v \in V.$$
2. *The length of a shortest bilinear computation for ϕ is called the bilinear complexity or the rank of ϕ and is denoted by $R(\phi)$.*
3. *If A is a finite dimensional associative k-algebra with unity, then the rank of A is defined as the rank of the multiplication map of A, which is a bilinear map $A \times A \to A$, and is denoted by $R(A)$.*

If we want to emphasize the underlying ground field k, we will sometimes write $R_k(\phi)$ resp. $R_k(A)$ instead of $R(\phi)$ resp. $R(A)$. In the remainder of this work, the term "algebra" always means a finite dimensional associative algebra with unity. For a k-algebra B, let $\mathrm{M}_n(B)$ denote the k-algebra of all $n \times n$–matrices with entries from B (with the usual definition of addition and multiplication). Using the language from above, the rank of $n \times n$–matrix multiplication over some field k is then denoted by $R(\mathrm{M}_n(k))$.

Enourmous effort has been spent on obtaining upper bounds for $R(\mathrm{M}_n(k))$. The current "world record" is $R(\mathrm{M}_n(k)) = O(n^{2.38})$, see [7]. In this paper, we are concerned with lower bounds. The main result is the following new lower bound.

Theorem 2 *Let k be an arbitrary field, and let $\mathrm{M}_n(k)$ be the k-algebra of $n \times n$–matrices with entries from k. Then*

$$R(\mathrm{M}_n(k)) \geq \tfrac{5}{2}n^2 - 3n. \tag{1}$$

For the special case $k = GF(2)$, Bshouty [5] has already obtained a similar bound (with the slightly worse lower order term $-\frac{1}{2}n\log n - O(n)$) using methods from coding theory. He also proved this bound for the multiplicative complexity.

Let us give a brief outline of the history of lower bounds for the rank of matrix multiplication: In 1978, Brockett and Dobkin [4] proved the lower bound

$$R(\mathrm{M}_n(k)) \geq 2n^2 - 1,$$

and in the same year, Lafon and Winograd [12] extended this result to multiplicative complexity. Three years later, Alder and Strassen [1] unified most of the lower bounds known at that time, including the last one, in a single theorem: For any finite dimensional associative k-algebra A

$$R(A) \geq 2\dim A - t,$$

where t is the number of maximal two-sided ideals in A. In fact, this bound even holds for the multiplicative complexity. Recently, some progress has been made (see [3]), namely

$$R(\mathrm{M}_n(k)) \geq 2n^2 + n - 3. \tag{2}$$

Once more, this bound also holds for the multiplicative complexity.

With the methods developed in this work, lower bounds similar to (1) can be obtained for various classes of algebras. For any noncommutative division algebra D, we here prove the lower bound $R(D) \geq \frac{5}{2}\dim D - 3$. As a third example, we prove the lower bound of $\frac{5}{4}n^2 - \frac{5}{4}n + 1$ for the rank of the multiplication of upper triangular matrices of size $n \times n$. Further examples can be found in [2].

We conclude this section with an outline of the proof. The main ingredient that allows us to achieve the improved results is Lemma 3 of the next section. An immediate application is the lower bound $R(D) \geq \frac{5}{2}\dim D - 3$ for any noncommutative division algebra D. While this result is interesting on its own, we will also see from the proof how to apply our method to matrix multiplication and what additional problems we have to solve. The main problem is to find two matrices $a, b \in \mathrm{M}_n(k)$ such that $ab - ba$ has full rank, or at least "large" rank. (Here, we mean the usual rank of a matrix which should not be confused with the rank of a bilinear map.) Without any restrictions, this task is quite easy. For example, we could choose a matrix a_0 as a diagonal matrix with pairwise distinct entries in the diagonal and b_0 as the matrix that has ones on the first superdiagonal and in the position $(n, 1)$ and zeros elsewhere. An easy calculation shows that $a_0b_0 - b_0a_0$ is invertible, that is, has full rank. But to put Lemma 3 into effective use, we have to deal with some restrictions: An adversary gives us an arbitrary basis $x_1, \dots, x_N$ of $\mathrm{M}_n(k)$, where $N = n^2$. Our task is to find a "small" subset $x_{i_1}, \dots, x_{i_t}$, such that the linear span of $x_{i_1}, \dots, x_{i_t}$ already contains matrices a, b with the above property. Here "small" means $t = o(n^2)$. The above a_0 and b_0 might not work, because the adversary might have given us a basis $x_1, \dots, x_N$ such that a_0 and b_0 are not contained in the linear span of any proper subset of $x_1, \dots, x_N$. Section 4 provides a solution of this problem.

2. Lower Bounds for the Rank of Noncommutative Algebras

Only a few standard methods are known for proving lower bounds in algebraic complexity theory. One of them is the so-called substitution method, due to Pan [13]. Strassen [15, Sec. 4] introduces another useful technique, which he uses to classify all 3-slice tensors of maximal border rank. Here, we combine these methods and apply them to associative algebras.

For an associative algebra A, let $[a, b] := ab - ba$ denote the Lie product of the two elements $a, b \in A$.

Lemma 3 *Let k be a field, A an associative k-algebra, $\dim A = N$, and let $\beta = (f_1, g_1, w_1, \dots, f_r, g_r, w_r)$ be a bilinear computation for A. If $1, a, b \in \bigcap_{\mu=1}^{m} \ker f_\mu$, then*

$$r \geq m + N + \tfrac{1}{2}\dim([a, b]A).$$

Proof. Consider the set of all g_ρ with $m + 1 \leq \rho \leq r$ and $f_\rho(1) \neq 0$. This set generates A^*, because

$$y \in \bigcap_{\substack{\rho=m+1,\\ f_\rho(1)\neq 0}}^{r} \ker g_\rho$$

implies

$$1 \cdot y = \sum_{\rho=1}^{r} f_\rho(1)g_\rho(y)w_\rho = \sum_{\substack{\rho=m+1,\\ f_\rho(1)\neq 0}}^{r} f_\rho(1)g_\rho(y)w_\rho = 0.$$

Therefore, we may assume w.l.o.g. that $g_{m+1}, \dots, g_{m+N}$ form a basis of A^* and $f_{m+\nu}(1) \neq 0$ for all $1 \leq \nu \leq N$. By appropriate scaling, we can achieve $f_{m+\nu}(1) = 1$ for all $1 \leq \nu \leq N$. Let $y_1, \dots, y_N$ be the dual basis of $g_{m+1}, \dots, g_{m+N}$. Then

$$y_\nu = 1 \cdot y_\nu = w_{m+\nu} + \sum_{\rho=m+N+1}^{r} f_\rho(1)g_\rho(y_\nu)w_\rho,$$

because $f_1(1) = \cdots = f_m(1) = 0$ and $g_{m+\nu'}(y_\nu) = 0$ for $1 \leq \nu' \leq N$ and $\nu' \neq \nu$, whence

$$w_{m+\nu} = y_\nu - \sum_{\rho=m+N+1}^{r} f_\rho(1)g_\rho(y_\nu)w_\rho. \tag{3}$$

Thus,

$$ay_\nu = f_{m+\nu}(a)w_{m+\nu} + \sum_{\rho=m+N+1}^{r} f_\rho(a)g_\rho(y_\nu)w_\rho \tag{4}$$

$$= f_{m+\nu}(a)y_\nu + \sum_{\rho=m+N+1}^{r} \left(f_\rho(a) - f_{m+\nu}(a)f_\rho(1)\right) g_\rho(y_\nu)w_\rho, \tag{5}$$

and in the same way,

$$by_\nu = f_{m+\nu}(b)w_{m+\nu} + \sum_{\rho=m+N+1}^{r} f_\rho(b)g_\rho(y_\nu)w_\rho \tag{6}$$

$$= f_{m+\nu}(b)y_\nu + \sum_{\rho=m+N+1}^{r} \left(f_\rho(b) - f_{m+\nu}(b)f_\rho(1)\right) g_\rho(y_\nu)w_\rho. \tag{7}$$

Multiplying Equation (5) with b from the left and using (6), we obtain

$$\begin{aligned} bay_\nu &= f_{m+\nu}(a)by_\nu \\ &\quad + \sum_{\rho=m+N+1}^{r} \left(f_\rho(a) - f_{m+\nu}(a)f_\rho(1)\right) g_\rho(y_\nu)bw_\rho \\ &= f_{m+\nu}(a)f_{m+\nu}(b)w_{m+\nu} \\ &\quad + \sum_{\rho=m+N+1}^{r} f_{m+\nu}(a)f_\rho(b)g_\rho(y_\nu)w_\rho \\ &\quad + \sum_{\rho=m+N+1}^{r} \left(f_\rho(a) - f_{m+\nu}(a)f_\rho(1)\right) g_\rho(y_\nu)bw_\rho. \end{aligned} \tag{8}$$

In the same manner, by multiplying (7) from the left with a, we get

$$\begin{aligned} aby_\nu &= f_{m+\nu}(b)f_{m+\nu}(a)w_{m+\nu} \\ &\quad + \sum_{\rho=m+N+1}^{r} f_{m+\nu}(b)f_\rho(a)g_\rho(y_\nu)w_\rho \\ &\quad + \sum_{\rho=m+N+1}^{r} \left(f_\rho(b) - f_{m+\nu}(b)f_\rho(1)\right) g_\rho(y_\nu)aw_\rho. \end{aligned} \tag{9}$$

Subtracting Equation (8) from (9) finally yields

$$\begin{aligned} [a,b]y_\nu = \sum_{\rho=m+N+1}^{r} \Big(& f_{m+\nu}(b)f_\rho(a)g_\rho(y_\nu)w_\rho \\ & + \left(f_\rho(b) - f_{m+\nu}(b)f_\rho(1)\right) g_\rho(y_\nu)aw_\rho \\ & - f_{m+\nu}(a)f_\rho(b)g_\rho(y_\nu)w_\rho \\ & - \left(f_\rho(a) - f_{m+\nu}(a)f_\rho(1)\right) g_\rho(y_\nu)bw_\rho \Big). \end{aligned}$$

The ρ-th summand in the above sum is contained in the linear span of the following three vectors

$$\begin{array}{rclll} v_{\rho,1} &=& f_\rho(a)w_\rho & - f_\rho(1)aw_\rho, & \\ v_{\rho,2} &=& -f_\rho(b)w_\rho & & + f_\rho(1)bw_\rho, \\ v_{\rho,3} &=& & f_\rho(b)aw_\rho & - f_\rho(a)bw_\rho \end{array}$$

for all ν with $1 \le \nu \le N$. Since the determinant of the coefficients with respect to w_ρ, aw_ρ, and bw_ρ vanishes, these three vectors are linearly dependent. (This elegant argument is due to D. Lauer.) Thus, the ρ-th summand is even contained in a vector space of dimension at most two, and this vector space is independent of ν. Since $y_1, \dots, y_N$ form a basis of A, $[a,b]y_1, \dots, [a,b]y_N$ generate $[a,b]A$. Therefore, $[a,b]A$ is contained in the sum of $r-m-N$ vector spaces of dimension two or less, yielding

$$\dim([a,b]A) \le 2(r-m-N),$$

which proves the lemma. ■

3. Division Algebras

An immediate application of Lemma 3 is a lower bound for the rank of noncommutative division algebras. An algebra is called a division algebra, if all nonzero elements of the algebra are invertible.

Before studying this case, let us first briefly recall the so-called *sandwiching*: Let $\beta = (f_1, g_1, w_1, \dots, f_r, g_r, w_r)$ be a bilinear computation for an algebra A, and let a, b, c be invertible elements of A. We have

$$\begin{aligned} xy &= a^{-1}(axb^{-1})(byc^{-1})c \\ &= \sum_{\rho=1}^{r} f_\rho(axb^{-1})g_\rho(byc^{-1})a^{-1}w_\rho c \end{aligned}$$

for all $x, y \in A$. Therefore $\tilde\beta = (\tilde f_1, \tilde g_1, \tilde w_1, \dots, \tilde f_r, \tilde g_r, \tilde w_r)$ is a bilinear computation for A, where the linear forms $\tilde f_\rho$ and $\tilde g_\rho$ are defined by $\tilde f_\rho(x) = f_\rho(axb^{-1})$ for all $x \in A$ resp. $\tilde g_\rho(y) = g_\rho(byc^{-1})$ for all $y \in A$, and $\tilde w_\rho = a^{-1}w_\rho c$. This defines an equivalence relation on the set of all computations of length r for A. Due to the shape of the above equation, the process of replacing β by $\tilde\beta$ is called sandwiching. For further background see [8].

Let D be a noncommutative division algebra over some field k, $\dim D = N$, and $\beta = (f_1, g_1, w_1, \dots, f_r, g_r, w_r)$ an optimal computation for D. W.l.o.g. we may assume after suitable permutation of the products that $f_1, \dots, f_N$ form a basis of D^*, because otherwise we would be able to find some $a \ne 0$ with $f_1(a) = \dots = f_r(a) = 0$. But then

$$a = a \cdot 1 = \sum_{\rho=1}^{r} f_\rho(a)g_\rho(1)w_\rho = 0,$$

which is a contradiction. Let $x_1, \ldots, x_N$ be the dual basis of $f_1, \ldots, f_N$. By sandwiching in a suitable way with x_N^{-1}, we can achieve $x_N = 1$. Since D is a noncommutative algebra, two of the elements $x_1, \ldots, x_{N-1}$ do not commute. After suitable renumbering, we may assume w.l.o.g. that these two elements are x_{N-2} and x_{N-1}. By Lemma 3 (with $m = N - 3$),

$$r \geq N - 3 + N + \tfrac{1}{2}\dim([x_{N-2}, x_{N-1}]A) = \tfrac{5}{2}N - 3,$$

because $[x_{N-2}, x_{N-1}]$ is nonzero and therefore invertible. As β is optimal, we get the following theorem.

Theorem 4 *Let D be a finite dimensional noncommutative division algebra over a field k. Then*

$$R(D) \geq \tfrac{5}{2}\dim D - 3.$$

The best lower bound known so far for the rank of a general division algebra is due to Baur (see [6, Theorem 17.27]): For any division algebra D over a field k there is an element $a \in D$, such that

$$R(D) \geq 2\dim D - 2 + \dim D / \dim k(a).$$

An example for noncommutative division algebras are cyclic division algebras: Let $\omega_n \in \mathbb{C}$ be a primitive n-th root of unity, then there is a noncommutative central division algebra over $\mathbb{Q}(\omega_n)$ of dimension n^2 (see for instance [16, §94]). Since cyclic division algebras of dimension n^2 contain subfields of dimension n, Baur's bound yields $R(D) \geq 2n^2 + n - 2$ for a cyclic division algebra D of dimension n^2, while we get $R(D) \geq \frac{5}{2}n^2 - 3$ from our new bound of Theorem 4, which is an improvement for $n \geq 3$.

4. Some Lie Theory

The proof of Theorem 4 shows that in order to put Lemma 3 into effective use in the case of the algebra $\mathrm{M}_n(k)$, we should be able to solve the following problem: Given a basis $x_1, \ldots, x_N$ of $\mathrm{M}_n(k)$, find a "small" subset $x_{i_1}, \ldots, x_{i_t}$ of the basis elements whose linear span contains elements a, b such that $\mathrm{rk}[a, b]$ is "large". Here, for a matrix $y \in \mathrm{M}_n(k)$, $\mathrm{rk}\, y$ denotes the rank of y. It turns out that we can achieve $t \leq 2n$ and $\mathrm{rk}[a, b] = n$, that is, $[a, b]$ has full rank. This is shown in the present section.

If A is an associative algebra and $a \in A$, then in the following $\mathrm{ad}\, a := [a, .\,]$ denotes the vector space endomorphism defined by $x \mapsto [a, x]$.

Lemma 5 *Let k be an infinite field, and let $v_1, \ldots, v_q \in \mathrm{M}_n(k)$ be such that the linear span of $v_1, \ldots, v_q$ contains an invertible matrix. Then there are $s \leq n$ and indices $i_1, \ldots, i_s$, such that already the linear span of $v_{i_1}, \ldots, v_{i_s}$ contains an invertible matrix.*

Proof. There are scalars $\alpha_1, \ldots, \alpha_q \in k$ such that $w = \alpha_1 v_1 + \cdots + \alpha_q v_q$ is invertible. Let $X_1, \ldots, X_q$ be indeterminates over k and consider the form

$$D(X_1, \ldots, X_q) = \det(X_1 v_1 + \cdots + X_q v_q)$$

of degree n in $X_1, \ldots, X_q$. The form D is nonzero, since $D(\alpha_1, \ldots, \alpha_q) = \det(w) \neq 0$. Let $\gamma X_{i_1}^{e_1} \ldots X_{i_s}^{e_s}$ be a monomial of D with $\gamma \neq 0$ and $e_\sigma > 0$ for all $1 \leq \sigma \leq s$. Since D has degree n, we have $s \leq n$. The form

$$D_0(X_{i_1}, \ldots, X_{i_s}) = \det(X_{i_1} v_{i_1} + \cdots + X_{i_s} v_{i_s}),$$

which is the homomorphic image of D under the substitution $X_j \mapsto 0$ for all $j \notin \{i_1, \ldots, i_s\}$, is also nonzero, because $\gamma X_{i_1}^{e_1} \ldots X_{i_s}^{e_s}$ is also a monomial of D_0. Since k is infinite, there are $\beta_1, \ldots, \beta_s \in k$ with $D_0(\beta_1, \ldots, \beta_s) \neq 0$, and therefore $\beta_1 v_{i_1} + \cdots + \beta_s v_{i_s}$ is invertible. ■

In the following, if V is a vector space and $v_1, \ldots, v_q$ are elements of V, then $\mathrm{lin}\,\{v_1, \ldots, v_q\}$ denotes the linear span of $v_1, \ldots, v_q$.

Lemma 6 *Let k be an infinite field, let $n \geq 2$, and let $x_1, \ldots, x_N$ be a basis of $\mathrm{M}_n(k)$, where $N = n^2$. Then there are $t \leq 2n$, indices $i_1, \ldots, i_t$, and $a, b \in \mathrm{lin}\,\{x_{i_1}, \ldots, x_{i_t}\}$, such that $[a, b]$ is invertible.*

Proof. Let $v = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \in \mathrm{M}_n(k)$ with pairwise distinct λ_ν, and let $e_{i,j} \in \mathrm{M}_n(k)$ be the matrix that has a one in position (i, j) and zeros in all other positions. A straight forward computation shows that for all i the matrix $e_{i,i}$ is an eigenvector of $\mathrm{ad}\, v$ associated with the eigenvalue 0 and that for all $i \neq j$ the matrix $e_{i,j}$ is an eigenvector of $\mathrm{ad}\, v$ associated with the eigenvalue $\lambda_i - \lambda_j \neq 0$. Hence

$$\ker(\mathrm{ad}\, v) = \mathrm{lin}\,\{e_{i,i} \mid 1 \leq i \leq n\} \text{ and}$$
$$I := \mathrm{im}(\mathrm{ad}\, v) = \mathrm{lin}\,\{e_{i,j} \mid i \neq j\}\,.$$

After suitable renumbering, we may assume w.l.o.g. that $\mathrm{ad}\, v(x_1), \ldots, \mathrm{ad}\, v(x_{N-n})$ form a basis of I. Since I contains an invertible matrix ($e_{1,2} + \cdots + e_{n-1,n} + e_{n,1}$ is in I, for instance), after another renumbering, we may assume by Lemma 5, that $\mathrm{lin}\,\{\mathrm{ad}\, v(x_1), \ldots, \mathrm{ad}\, v(x_s)\}$ contains an invertible matrix, where $s \leq n$. Thus there are scalars $\alpha_1, \ldots, \alpha_s \in k$, such that

$$\alpha_1[v, x_1] + \cdots + \alpha_s[v, x_s] = [v, \alpha_1 x_1 + \cdots + \alpha_s x_s]$$

is invertible. As $[u, w] = -[w, u]$ for all $u, w \in \mathrm{M}_n(k)$, $[a, v]$ is also invertible, where $a = \alpha_1 x_1 + \cdots + \alpha_s x_s$. Therefore $\mathrm{im}(\mathrm{ad}\, a)$ contains an invertible matrix. Since the matrices $\mathrm{ad}\, a(x_1), \ldots, \mathrm{ad}\, a(x_N)$ span $\mathrm{im}(\mathrm{ad}\, a)$, there are, by Lemma 5, indices $j_1, \ldots, j_{s'}$, where $s' \leq n$, such that

$\mathrm{lin}\left\{\mathrm{ad}\, a(x_{j_1}), \ldots, \mathrm{ad}\, a(x_{j_{s'}})\right\}$ contains an invertible matrix. In other words, there are scalars $\beta_1, \ldots, \beta_{s'} \in k$, such that

$$\beta_1[a, x_{j_1}] + \cdots + \beta_{s'}[a, x_{j_{s'}}] = [a, \beta_1 x_{j_1} + \cdots + \beta_{s'} x_{j_{s'}}]$$

is invertible. Therefore, a and $b = \beta_1 x_{j_1} + \cdots + \beta_{s'} x_{j_{s'}}$ are matrices with the desired property. Since $s + s' \leq 2n$, the lemma is proven. ■

5. Matrix Multiplication

In the proof of Lemma 6, we use the assumption that the ground field k is infinite. Since we are considering lower bounds, this is no restriction: The definition of rank immediately yields for every r a system of polynomial equations over k which has a solution if and only if $R_k(\mathrm{M}_n(k)) \leq r$. Enlarging the ground field can only produce more solutions, hence lower bounds over some ground field hold over all of its subfields. Speaking more algebraically, if $K \supset k$ is an extension field, then for all k-algebras A

$$R_K(A \otimes K) \leq R_k(A), \tag{10}$$

where $A \otimes K$ denotes the tensor product of A and K, which is a K-algebra with $\dim_K(A \otimes K) = \dim_k A$. (10) is easily seen as follows: If $\beta = (f_1, w_1, g_1, \ldots, f_r, w_r, g_r)$ is an optimal bilinear computation over k for A, then $\bar{\beta} = (\bar{f}_1, \bar{g}_1, \bar{w}_1, \ldots, \bar{f}_r, \bar{g}_r, \bar{w}_r)$ is a bilinear computation over K for the K-algebra $A \otimes K$, where $\bar{f}_\rho(x \otimes \xi) = \xi f_\rho(x)$, $\bar{g}_\rho(y \otimes \eta) = \eta g_\rho(y)$, and $\bar{w}_\rho = w_\rho \otimes 1$, because

$$\begin{aligned}\sum_{\rho=1}^{r} \bar{f}_\rho(x \otimes \xi)\bar{g}_\rho(y \otimes \eta)(\bar{w}_\rho \otimes 1)& \\ &= \xi\eta \left(\sum_{\rho=1}^{r} f_\rho(x) g_\rho(y) w_\rho\right) \otimes 1 \\ &= xy \otimes \xi\eta\end{aligned}$$

for all $x, y \in A$, $\xi, \eta \in K$. Since $\mathrm{M}_n(k) \otimes K = \mathrm{M}_n(K)$, we may assume w.l.o.g. in the following that the ground field is infinite. Now we are able to prove our main theorem.

Theorem 2 *Let k be an arbitrary field, and let $\mathrm{M}_n(k)$ be the k-algebra of $n \times n$–matrices with entries from k. Then*

$$R(\mathrm{M}_n(k)) \geq \tfrac{5}{2}n^2 - 3n.$$

Proof. W.l.o.g. we may assume that k is infinite. Let $N = \dim \mathrm{M}_n(k)$. Let $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ be an optimal bilinear computation for $\mathrm{M}_n(k)$. After suitable renumbering, we may assume that $f_1, \ldots, f_N$ form a basis of $\mathrm{M}_n(k)^*$. Let $x_1, \ldots, x_N$ denote the dual basis of $f_1, \ldots, f_N$. By Lemma 5, there are indices $i_1, \ldots, i_s$, where $s \leq n$, such that $\mathrm{lin}\{x_{i_1}, \ldots, x_{i_s}\}$ contains an invertible matrix e. By appropriate sandwiching, we can achieve $e = 1$. By Lemma 6, there are $t \leq 2n$, indices $j_1, \ldots, j_t$, and $a, b \in \mathrm{lin}\{x_{j_1}, \ldots, x_{j_t}\}$, such that $[a, b]$ is invertible. After another suitable renumbering, we may assume that $N - 3n + 1 \leq i_1, \ldots, i_s, j_1, \ldots, j_t \leq N$, since the number of indices $s+t$ is at most $3n$. Thus, by Lemma 3 (with $m = N - 3n$),

$$r \geq 2N - 3n + \tfrac{1}{2}\dim([a, b]\mathrm{M}_n(k)) = \tfrac{5}{2}n^2 - 3n. \quad ■$$

Compared with the bound (2), our new bound is an improvement for $n \geq 8$.

Another nice application of Lemma 3 is the multiplication of upper triangular matrices. Let $\mathrm{T}_n(k)$ denote the k-algebra of upper triangular matrices of size $n \times n$ with entries from k. Applying the Alder-Strassen bound to $\mathrm{T}_n(k)$, we obtain $R(\mathrm{T}_n(k)) \geq 2\dim \mathrm{T}_n(k) - n$. Heintz and Morgenstern [10] show that strict inequality holds for $n > 2$. The following theorem provides a new bound, which improves the bound of Heintz and Morgenstern for $n \geq 6$.

Theorem 7 *Let k be a field, and let $\mathrm{T}_n(k)$ be the k-algebra of upper triangular matrices of size $n \times n$ with entries from k. Then*

$$R(\mathrm{T}_n(k)) \geq \tfrac{5}{2}\dim \mathrm{T}_n(k) - \tfrac{5}{2}n + 1.$$

Proof. Let $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ be an optimal computation for $\mathrm{T}_n(k)$. Let $N = \dim \mathrm{T}_n(k)$. Let $I \subset \mathrm{T}_n(k)$ be the two-sided ideal of all matrices that have a zero in each position of the main diagonal. Let $J = I^2$. Then J is the two-sided ideal of all matrices that have a zero in each position of the main and of the first superdiagonal. By suitable renumbering, we may assume that

$$J \oplus \bigcap_{\nu=1}^{N-2n+1} \ker f_\nu = \mathrm{T}_n(k).$$

Let $X = \bigcap_{\nu=1}^{N-2n+1} \ker f_\nu$. Then X contains an invertible matrix e. (There is some $c \in J$ such that $1 + c \in X$, for instance.) By suitable sandwiching, we may assume $e = 1$. Let $a = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ with pairwise distinct α_ν, and let $b \in \mathrm{T}_n(k)$ be the matrix that has a one in each position of the first superdiagonal and zeros elsewhere. There are $a', b' \in J$ such that $a + a', b + b' \in X$. Applying Lemma 3 (with $m = N - 2n + 1$), we obtain

$$r \geq 2N - 2n + 1 + \tfrac{1}{2}\dim([a + a', b + b']\mathrm{T}_n(k)).$$

Since J is a two-sided ideal, there is some $c \in J$ fulfilling $[a + a', b + b'] = [a, b] + c$. An easy calculation shows that

$[a, b] \in I$ and that all entries of $[a, b]$ in the first superdiagonal are nonzero. Since $c \in J$, the same holds for $[a, b] + c$. Therefore $([a, b] + c)\mathrm{T}_n(k) = I$, whence

$$\dim([a + a', b + b']\mathrm{T}_n(k)) = \dim \mathrm{T}_n(k) - n. \quad \blacksquare$$

There are more applications of Lemma 3, for example simple algebras. A k-algebra A is called simple, if it contains no proper two-sided ideals other than the zero ideal. If A is a simple algebra, then, by Wedderburn's theorem, there is an (unique) division algebra D over k and an (unique) integer $n \geq 1$, such that A is isomorphic to $\mathrm{M}_n(D)$. Refining the methods of the preceding section, we obtain $R(A) \geq \frac{5}{2} \dim A - 3n$, if $n \geq 2$.

The bound from Theorem 2 also holds in the following, more general, case. For a k-algebra A, let $\mathrm{rad}\, A$ denote its radical (see e.g. [16, §96]). Then $A/\mathrm{rad}\, A$ is semisimple and, by Wedderburn's theorem, isomorphic to a finite direct product of algebras $\mathrm{M}_{n_\tau}(D_\tau)$, $1 \leq \tau \leq t$, where the D_τ are k-division algebras. If now each factor is noncommutative, that is, $n_\tau \geq 2$ or the algebra D_τ is noncommutative, then $R(A) \geq \frac{5}{2} \dim A - 3n$, where $n = n_1 + \cdots + n_t$. The proof of this generalization is beyond the scope of the present work, see the forthcoming [2].

Acknowledgments

I would especially like to thank A. Schönhage for introducing me to this topic, for the abundance of fruitful discussions, and for his valuable suggestions. I am grateful for helpful comments from and discussions with P. Kirrinnis and D. Lauer. I would also like to thank the anonymous referees for their useful comments.

This work was supported through funds provided by the Deutsche Forschungsgemeinschaft under Grant No. Scho 415/3-1.

References

[1] A. Alder and V. Strassen. On the algorithmic complexity of associative algebras. *Theoret. Comput. Sci.*, 15:201–211, 1981.

[2] M. Bläser. *Untere Schranken für den Rang assoziativer Algebren*. Dissertation, Universität Bonn, in preparation.

[3] M. Bläser. Lower bounds for the multiplicative complexity of matrix multiplication. *Comput. Complexity*, to appear.

[4] R. W. Brockett and D. Dobkin. On the optimal evaluation of a set of bilinear forms. *Lin. Alg. Appl.*, 19:207–235, 1978.

[5] N. H. Bshouty. A lower bound for matrix multiplication. *SIAM J. Comput.*, 18:759–765, 1989.

[6] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic Complexity Theory*. Springer, 1997.

[7] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progression. *J. Symb. Comp.*, 9:251–280, 1990.

[8] H. F. de Groote. On the varieties of optimal algorithms for the computation of bilinear mappings: The isotropy group of a bilinear mapping. *Theoret. Comput. Sci.*, 7:1–24, 1978.

[9] H. F. de Groote. *Lectures on the Complexity of Bilinear Problems*. LNCS 245. Springer, 1986.

[10] J. Heintz and J. Morgenstern. On associative algebras of minimal rank. In *Proceedings of the AAECC 2*, LNCS 228, pages 1–24. Springer, 1986.

[11] N. Jacobson. *Lie Algebras*. Interscience Publishers, third edition, 1966.

[12] J. C. Lafon and S. Winograd. A lower bound for the multiplicative complexity of the product of two matrices. Technical report, Centre de Calcul de L'Esplanade, U.E.R. de Mathematique, Univ. Louis Pasteur, Strasbourg, France, 1978.

[13] V. Pan. Methods for computing values of polynomials. *Russ. Math. Surv.*, 21:105–136, 1966.

[14] V. Strassen. Vermeidung von Divisionen. *J. Reine Angew. Math.*, 264:184–202, 1973.

[15] V. Strassen. Rank and optimal computation of generic tensors. *Lin. Alg. Appl.*, 52/53:645–685, 1983.

[16] B. L. van der Waerden. *Algebra II*. Springer, fifth edition, 1967.