Improvements of the Alder–Strassen Bound: Algebras with Nonzero Radical

Markus Bläser

Institut für Theoretische Informatik, Med. Universität zu Lübeck
Wallstr. 40, 23560 Lübeck, Germany
blaeser@tcs.mu-luebeck.de

Abstract. Let $C(A)$ denote the multiplicative complexity of a finite dimensional associative k-algebra A.
For algebras A with nonzero radical $\operatorname{rad} A$, we exhibit several lower bound techniques for $C(A)$ that yield bounds significantly above the Alder–Strassen bound. In particular, we prove that the multiplicative complexity of the multiplication in the algebras $k[X_1, \ldots, X_n]/\, I_{d+1}(X_1, \ldots, X_n)$ is bounded from below by $3 \cdot \binom{n+d}{n} - \binom{n+\lceil d/2 \rceil}{n} - \binom{n+\lfloor d/2 \rfloor}{n}$, where $I_d(X_1, \ldots, X_n)$ denotes the ideal generated by all monomials of degree d in $X_1, \ldots, X_n$. Furthermore, we show the lower bound $C(\mathrm{T}_n(k)) \geq (2\frac{1}{8} - o(1)) \dim \mathrm{T}_n(k)$ for the multiplication of upper triangular matrices.

1 Introduction

A fundamental problem in algebraic complexity theory is the question about the costs of multiplication, say of matrices, triangular matrices, polynomials, or power series, just to mention a few. To be more specific, let A be a finite dimensional associative k-algebra with unity 1. By fixing a basis of A, say $v_1, \ldots, v_N$, we can define a set of bilinear forms corresponding to the multiplication in A. If $v_\mu v_\nu = \sum_{\kappa=1}^{N} \alpha_{\mu,\nu}^{(\kappa)} v_\kappa$ for $1 \leq \mu, \nu \leq N$ with *structural constants* $\alpha_{\mu,\nu}^{(\kappa)} \in k$, then these constants and the identity

$$\left(\sum_{\mu=1}^{N} X_\mu v_\mu \right) \left(\sum_{\nu=1}^{N} Y_\nu v_\nu \right) = \sum_{\kappa=1}^{N} b_\kappa(X, Y) v_\kappa$$

define the desired bilinear forms $b_1, \ldots, b_N$. The *multiplicative complexity* of $b_1, \ldots, b_N$ is the smallest number of *essential* multiplications and divisions necessary and sufficient to compute $b_1, \ldots, b_N$ from the indeterminates $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$.

According to Strassen [15], we may reformulate the problem over infinite fields as follows: the multiplicative complexity of $b_1, \ldots, b_N$ is the smallest number ℓ of products $p_\lambda = u_\lambda(X_i, Y_j) \cdot v_\lambda(X_i, Y_j)$ with linear forms u_λ and v_λ in the X_i and Y_j such that the linear span of $p_1, \ldots, p_\ell$ contains $b_1, \ldots, b_N$. (The restriction to infinite fields is not critical for this work, since we are concerned with lower bounds.) From this characterization, it follows that the multiplicative complexity of $b_1, \ldots, b_N$ does not depend on the choice of $v_1, \ldots, v_N$, thus we may speak about the multiplicative complexity of (the multiplication in) A. For a modern introduction to algebraic complexity theory, we recommend [8].

A fundamental lower bound for the multiplicative complexity is the so-called Alder–Strassen bound (1) (see Section 1.2). Recently, this bound has been improved for a large class of semisimple[1] algebras [6] as shown in (2). The main contributions of this work are improvements of the Alder–Strassen bound for algebras with nonzero radical, like upper triangular matrices.

Closely related to the multiplicative complexity is the *bilinear complexity* (or *rank*). Here the products $p_\lambda = u_\lambda(X_i) \cdot v_\lambda(Y_j)$ are bilinear products, that is, products of linear forms u_λ in the X_i and linear forms v_λ in the Y_j. (Note that $b_1, \dots, b_N$ are bilinear forms.) The multiplicative complexity is clearly a lower bound for the bilinear complexity and it is easy to show that twice the multiplicative complexity is an upper bound for the bilinear complexity (see e.g. [8, Eq. 14.8]). Therefore, we usually want to derive upper bounds for the bilinear complexity and lower bounds for the multiplicative complexity.

While the difference between multiplicative and bilinear complexity seems to be minor at a first glance, it is much harder to cope with the multiplicative complexity when dealing with lower bounds. The main reason is the fact that the bilinear complexity of a tensor of a bilinear map (see below for a definition) is invariant under permutations whereas the multiplicative complexity might not, see also [8, Chap. 14.2] for a further discussion. For instance, if we consider bilinear complexity, then the bound (2) holds also for any algebra with nonzero radical, provided that the semisimple quotient algebra $A/\operatorname{rad} A$ fulfils the corresponding premises [4, 5]. On the other hand, there are examples given in [6, Sect. 6] that show that the novel methods from [6] for the multiplicative complexity do not apply to algebras with nonzero radical.

1.1 Model of Computation

For proving lower bounds, a coordinate-free definition of multiplicative complexity is often more appropriate than the one given above, see e.g. [8, Chap. 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, and W finite dimensional vector spaces over k, and $\phi : U \times V \to W$ a bilinear map.*

1. *A sequence $\beta = (f_1, g_1, w_1, \dots, f_\ell, g_\ell, w_\ell)$ with $f_\lambda, g_\lambda \in (U \times V)^*$ and $w_\lambda \in W$ is called a quadratic computation for ϕ over k of length ℓ if*

$$\phi(u, v) = \sum_{\lambda=1}^{\ell} f_\lambda(u, v) g_\lambda(u, v) w_\lambda \qquad \textit{for all } u \in U, v \in V.$$

[1] For a finite dimensional associative k-algebra A with unity, the *radical* $\operatorname{rad} A$ is the intersection of all maximal twosided ideals of A. An algebra A is called *semisimple* if $\operatorname{rad} A = \{0\}$, see [9, 14] for more details.

2. *The length of a shortest quadratic computation for ϕ is called the multiplicative complexity of ϕ and is denoted by $C(\phi)$.*
3. *If A is a finite dimensional associative k-algebra with unity, then the multiplicative complexity of A is defined as the multiplicative complexity of the multiplication map of A, which is a bilinear map $A \times A \to A$, and is denoted by $C(A)$.*

If we require that $f_\lambda \in U^*$ and $g_\lambda \in V^*$ in the above Definition 1, we get *bilinear computations* and *bilinear complexity* (also called *rank*). We denote the bilinear complexity of a bilinear map ϕ by $R(\phi)$ and the bilinear complexity of an associative algebra A by $R(A)$. We have $C(\phi) \leq R(\phi) \leq 2 \cdot C(\phi)$ for any bilinear map ϕ. Except for trivial cases, the second inequality is always strict, see [12].

1.2 Previous Results

The best general lower bound for the multiplicative complexity of an associative algebra A is due to Alder and Strassen [1], they show

$$C(A) \geq 2 \dim A - t\,, \tag{1}$$

where t is the number of maximal twosided ideals in A. This has recently been improved for a large class of semisimple algebras by Bläser [6]: if A is semisimple and if in the decomposition of $A \cong A_1 \times \cdots \times A_t$ into simple factors $A_\tau = {D_\tau}^{n_\tau \times n_\tau}$ with division algebra D_τ, each A_τ is noncommutative, then

$$C(A) \geq \tfrac{5}{2} \dim A - 3(n_1 + \cdots + n_t). \tag{2}$$

Specifically, the multiplicative complexity of $n \times n$–matrix multiplication is at least $\frac{5}{2}n^2 - 3n$. While the lower bound (2) also holds for algebras with nonzero radical in the case of bilinear complexity [4, 5] provided that $A/\operatorname{rad} A$ fulfils the above condition, examples are presented in [6, Sec. 6] that show that those methods do not transfer to the multiplicative complexity for algebras with nonzero radical.

1.3 New Results

The starting point of our work is the above observation that the results from [4, 5] for the bilinear complexity do not transfer to the multiplicative complexity if A has nonzero radical. As our main contribution, we improve the Alder-Strassen bound (1) for various classes of algebras with nonzero radical.

As our first main result, we obtain the lower bound

$$C(A) \geq \dim A + \dim(\operatorname{rad} A)^m + \dim(\operatorname{rad} A)^n - \dim(\operatorname{rad} A)^{n+m-1}.$$

for any $n, m > 0$ (Theorem 1). This bound works very well for algebras with "growing" radical, that means, $2\dim(\operatorname{rad} A)^m$ is much larger than $\dim \operatorname{rad} A$ where m is the smallest natural number such that $(\operatorname{rad} A)^{2m-1} = \{0\}$. As an example, we apply this bound to the algebras $k[X_1, \ldots, X_n]/\, I_{d+1}(X_1, \ldots, X_n)$ where $I_d(X_1, \ldots, X_n)$ denotes the ideal generated by all monomials of degree d in $X_1, \ldots, X_n$. Furthermore, we obtain a sequence of explicitly given algebras A_n with $C(A_n) \geq (3 - o(1)) \dim A_n$. To our knowledge, this is the best lower bound known for a concrete algebra. (For existential results of algebras of high complexity, see [7].)

More complicated is the case of algebras with "nongrowing" radical, the most important one is probably the algebra of upper triangular $n \times n$–matrices $\mathrm{T}_n(k)$. We deal with this in Section 5. Our second main result is Lemma 7. This lemma provides a way to prove lower bounds above the Alder Strassen bound for algebras A with "nongrowing" radical. Here, $C(A)$ is estimated from below by the multiplicative complexity of a bilinear map ψ obtained from the multiplication in A by restricting to some subspaces. After that, $C(\psi)$ can be estimated using techniques introduced in [6]. In Section 6, we apply this method to the algebra of upper triangular matrices and get the lower bound

$$C(\mathrm{T}_n(k)) \geq (2\tfrac{1}{8} - o(1)) \dim \mathrm{T}_n(k).$$

This is the first bound for $\mathrm{T}_n(k)$ significantly above the Alder–Strassen bound. Prior to this, we only knew that the Alder–Strassen bound could be improved by an additive amount of one for $n \geq 3$ [11], that is, $\mathrm{T}_n(k)$ is not an algebra of minimal rank.

2 Preliminaries

For the reader's convenience, we compile some preliminaries to which we will refer frequently in the subsequent sections. In the first part of this section, we present an alternative characterization of multiplicative complexity, the so-called "tensorial notion" and state the relevant results. In the second part, we briefly review the lower bound techniques used by Alder and Strassen to prove their lower bound. These techniques basically are sophisticated refinements of the substitution method due to Pan [13].

2.1 Characterizations of Multiplicative Complexity

In the previous section, we have introduced the multiplicative complexity of a bilinear map in terms of computations. A second useful characterization of multiplicative complexity is the so-called "tensorial notion" (see [8, Chap. 14.4] for the bilinear complexity). With a bilinear map $\phi : U \times V \to W$, we may associate a *coordinate tensor* (or tensor for short) which is basically a "three-dimensional matrix": we fix bases $u_1, \ldots, u_m$ of U, $v_1, \ldots, v_n$ of V, and $w_1, \ldots, w_p$ of W. There are unique scalars $t_{\mu,\nu,\rho} \in k$ such that $\phi(u_\mu, v_\nu) = \sum_{\rho=1}^{p} t_{\mu,\nu,\rho} w_\rho$ for all $1 \leq \mu \leq m$, $1 \leq \nu \leq n$. Then $t = (t_{\mu,\nu,\rho}) \in k^{m \times n \times p}$ is the tensor of ϕ (with respect to the chosen bases). On the other hand, any given tensor also defines a bilinear map after choosing bases. We define the multiplicative complexity of the tensor t by $C(t) := C(\phi)$. In the same way, the bilinear complexity of t is $R(t) := R(\phi)$. (This is in both cases well-defined.)

With each tensor $t = (t_{\mu,\nu,\rho})$, we may associate three sets of matrices, the *slices* of t. The matrices $Q_\mu = (t_{\mu,\nu,\rho})_{1 \leq \nu \leq n, 1 \leq \rho \leq p} \in k^{n \times p}$ with $1 \leq \mu \leq m$ are called the 1-slices of t, the matrices $S_\nu = (t_{\mu,\nu,\rho})_{1 \leq \mu \leq m, 1 \leq \rho \leq p} \in k^{m \times p}$ with $1 \leq \nu \leq n$ the 2-slices, and finally $T_\rho = (t_{\mu,\nu,\rho})_{1 \leq \mu \leq m, 1 \leq \nu \leq n} \in k^{m \times n}$ with $1 \leq \rho \leq p$ are called the 3-slices of t. When dealing with bilinear complexity, it makes no difference which of the three sets of slices we consider. In the case of multiplicative complexity, however, the 3-slices play a distinguished role.

Lemma 1. *Let k be a field and t be a tensor with 3-slices $T_1, \dots, T_p \in k^{m \times n}$. Then $C(t) \le \ell$ if and only if there are (column) vectors $u_\lambda, v_\lambda \in k^{m+n}$ for $1 \le \lambda \le \ell$ such that with $P_\lambda := u_\lambda \cdot v_\lambda^\top \in k^{(m+n)\times(m+n)}$*

$$\begin{pmatrix} 0 & T_1 \\ T_1^\top & 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 & T_p \\ T_p^\top & 0 \end{pmatrix} \in \operatorname{lin}\{P_1 + P_1^\top, \dots, P_\ell + P_\ell^\top\}\,. \tag{3}$$

Here, $T^\top$ denotes the transpose of a matrix T and $\operatorname{lin}\{\dots\}$ denotes the linear span. A proof of this lemma is straight forward (see for instance [12, Thm. 3.2]).

If $T_1, \dots, T_p$ are the 3-slices of a tensor t, we will occasionally write $C(T_1, \dots, T_p)$ instead of $C(t)$ and $R(T_1, \dots, T_p)$ instead of $R(t)$. By multiplying (3) with

$$\begin{pmatrix} X & 0 \\ 0 & Y^\top \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} X^\top & 0 \\ 0 & Y \end{pmatrix}$$

from the left and right, respectively, it follows from Lemma 1 that if $X \in k^{m \times m}$ and $Y \in k^{n \times n}$ are invertible matrices, then

$$C(T_1, \dots, T_p) = C(X \cdot T_1 \cdot Y, \dots, X \cdot T_p \cdot Y)\,. \tag{4}$$

2.2 The Lower Bound Techniques of Alder and Strassen

Beside the original paper of Alder and Strassen, [10, Chap. IV.2] and [8, Chap. 17] are excellent treatments of the results of Alder and Strassen. We have taken the term "separate" and the extension lemma from there, but everything is also contained in the work of Alder and Strassen [1].

Definition 2. *Let U, V, W be vector spaces and $\beta = (f_1, g_1, w_1, \dots, f_\ell, g_\ell, w_\ell)$ be a quadratic computation for a bilinear map $\phi : U \times V \to W$. Let $U_1 \subseteq U$, $V_1 \subseteq V$, and $W_1 \subseteq W$ be subspaces. The computation β* separates *(U_1, V_1, W_1), if there is a set of indices $I \subseteq \{\lambda \mid w_\lambda \notin W_1\}$ such that after possibly exchanging some of the f_λ with the corresponding g_λ, we have $(U_1 \times V_1) \cap \bigcap_{i \in I} \ker f_i = \{0\}$.*

The latter condition is equivalent to the condition that $(f_i|_{U_1 \times V_1})_{i \in I}$ generate the dual space $(U_1 \times V_1)^*$.

If $\phi : U \times V \to W$ is a bilinear map and $U_1 \subseteq U$ and $V_1 \subseteq V$ are subspaces, then $(u + U_1, v + V_1) \mapsto \phi(u, v) + \tilde{W}$ defines a bilinear map $U/U_1 \times V/V_1 \to W/\tilde{W}$ where $\tilde{W} := \operatorname{lin}\{\phi(U_1, V)\} + \operatorname{lin}\{\phi(U, V_1)\}$. This map is called the *quotient* of ϕ by U_1 and V_1 and is denoted by $\phi/(U_1 \times V_1)$. We have the following lower bound. (See [8, Lem. 17.17] where also a proof is given.)

Lemma 2. *Let U, V, and W be vector spaces and $\beta = (f_1, g_1, w_1, \dots, f_\ell, g_\ell, w_\ell)$ be a quadratic computation for some bilinear map $\phi : U \times V \to W$. Let $U_1 \subseteq U$, $V_1 \subseteq V$, and $W_1 \subseteq W$ be subspaces such that β separates (U_1, V_1, W_1). Let π be an endomorphism of W such that $W_1 \subseteq \ker \pi$. Then*

$$\ell \ge C((\pi \circ \phi)/(U_1 \times V_1)) + \dim U_1 + \dim V_1 + \#\{\lambda \mid w_\lambda \in W_1\}.$$

If ϕ is the multiplication map of an associative algebra A and I is a twosided ideal of A, then $\phi/(I \times I)$ is the multiplication map of the quotient algebra A/I.

Corollary 1. *Let A be an algebra and $I \subseteq A$ a twosided ideal. Let β be a quadratic computation for A of length ℓ that separates $(I, I, \{0\})$. Then $\ell \geq C(A/I) + 2 \dim I$.*

To achieve good lower bounds by means of Lemma 2, one has to find an optimal bilinear computation which separates a "large" triple. An important tool to solve this task is the following "extension lemma", see [8, Lem. 17.18].

Lemma 3 (Alder and Strassen). *Let U, V, W be vector spaces and β be a quadratic computation for a bilinear map $\phi : U \times V \to W$. Let $U_1 \subseteq U_2 \subseteq U$, $V_1 \subseteq V$, and $W_1 \subseteq W$ be subspaces such that β separates (U_1, V_1, W_1). Then β separates also (U_2, V_1, W_1), or there is some $u \in U_2 \setminus U_1$ such that $\phi(u, V) \subseteq \operatorname{lin}\{\phi(U_1, V_1)\} + W_1$.*

In the course of their proof, Alder and Strassen first deal with the radical of an algebra A and then turn to the semisimple quotient algebra $A/\operatorname{rad} A$. The following lemma contains the first of the two important statements established by Alder and Strassen, see [1, Lem. 2] or [8, Prop. 17.20].

Lemma 4 (Alder and Strassen). *Let β be a quadratic computation for an associative algebra A. Then β separates $(\operatorname{rad} A, \operatorname{rad} A, \{0\})$.*

3 Lower Bounds: "Growing" Radicals

In this section, we develop a method that works very well if the radical of an algebra A "grows", more precisely, $2 \dim(\operatorname{rad} A)^m$ is much larger than $\dim \operatorname{rad} A$ where m is the smallest natural number such that $(\operatorname{rad} A)^{2m-1} = \{0\}$. This method has already been applied to the bilinear complexity, see [4, 5]. In contrast to the "hard" case (in Section 5), we do not loose anything when extending this method to the multiplicative complexity.

We build upon the following lower bound which is given in [2, Lem. 2].

Lemma 5. *Let A be an associative algebra. If $U, V \subseteq \operatorname{rad} A$ are vector spaces, then*

$$C(A) \geq \dim U + \dim V + \min_{\substack{X \oplus (U \times V) \\ = A \times A}} \{\dim \operatorname{lin}\{xx' \mid (x, x') \in X\}\}.$$

To put this lemma into effective use, we have to estimate the dimension of the subspaces $\operatorname{lin}\{xx' \mid (x, x') \in X\}$ in the above Lemma 5. This is done through the following lemma.

Lemma 6. *Let A be an associative algebra and $m, n > 0$ be natural numbers. For any vector space X such that $X \oplus ((\operatorname{rad} A)^m \times (\operatorname{rad} A)^n) = A \times A$,*

$$\operatorname{lin}\{xx' \mid (x, x') \in X\} + (\operatorname{rad} A)^{m+n-1} = A.$$

Proof. For the ease of notation, if $Y \subseteq A \times A$ is a vector space, let $\hat{Y}$ denote the vector space $\mathrm{lin}\{yy' \mid (y, y') \in Y\}$.

For $0 \leq \mu < m$ and $0 \leq \nu < n$, let $X_{\mu,\nu} \subseteq X$ be a vector space such that $X_{\mu,\nu} \oplus ((\mathrm{rad}\, A)^m \times (\mathrm{rad}\, A)^n) = (\mathrm{rad}\, A)^\mu \times (\mathrm{rad}\, A)^\nu$. Such an $X_{\mu,\nu}$ exists, because $X \oplus ((\mathrm{rad}\, A)^m \times (\mathrm{rad}\, A)^n) = A \times A$. As $X \cap ((\mathrm{rad}\, A)^m \times (\mathrm{rad}\, A)^n) = \{0\}$, it is also unique. Furthermore, $X_{\mu,\nu} \subseteq X_{\mu',\nu'}$ for $\mu' \leq \mu$ and $\nu' \leq \nu$.

For any $(u, v) \in (\mathrm{rad}\, A)^\mu \times (\mathrm{rad}\, A)^\nu$, there are $(a, b) \in (\mathrm{rad}\, A)^m \times (\mathrm{rad}\, A)^n$ such that $(u, v) + (a, b) \in X_{\mu,\nu}$. Thus $(u + a)(v + b) \in \hat{X}_{\mu,\nu}$. But $(u + a)(v + b) \in uv + (\mathrm{rad}\, A)^{\mu+\nu+1}$. Letting (u, v) run through all elements of $(\mathrm{rad}\, A)^\mu \times (\mathrm{rad}\, A)^\nu$, we get

$$\hat{X}_{\mu,\nu} + (\mathrm{rad}\, A)^{\mu+\nu+1} = (\mathrm{rad}\, A)^{\mu+\nu}. \tag{5}$$

We now prove by backward induction in $\mu + \nu$ that

$$\hat{X}_{\mu,\nu} + (\mathrm{rad}\, A)^{m+n-1} = (\mathrm{rad}\, A)^{\mu+\nu} \qquad \text{for all } \mu < m, \nu < n.$$

For $\mu = \nu = 0$, this is the claim of the lemma.

The induction start ($\mu = m - 1$, $\nu = n - 1$) follows directly from (5). For the induction step, let μ and ν be given such that $\mu + \nu < m + n - 2$. We assume that $\mu < m - 1$, the case $\nu < n - 1$ follows completely alike. By substituting the induction hypothesis $\hat{X}_{\mu+1,\nu} + (\mathrm{rad}\, A)^{m+n-1} = (\mathrm{rad}\, A)^{\mu+\nu+1}$ into (5), we obtain $\hat{X}_{\mu,\nu} + \hat{X}_{\mu+1,\nu} + (\mathrm{rad}\, A)^{m+n-1} = (\mathrm{rad}\, A)^{\mu+\nu}$. Now the claim follows from the fact that $\hat{U} \subseteq \hat{V}$ if $U \subseteq V$. $\square$

Combining the last lemma with Lemma 5 shows the following lower bound.

Theorem 1. *Let A be a an associative algebra. For all $m, n > 0$,*

$$C(A) \geq \dim A + \dim(\mathrm{rad}\, A)^m + \dim(\mathrm{rad}\, A)^n - \dim(\mathrm{rad}\, A)^{n+m-1}.$$

4 Multiplying Multivariate Power Series

For indeterminates $X_1, \ldots, X_n$, let $I_d(X_1, \ldots, X_n)$ denote the ideal generated by all monomials of degree d. Let $P_{n,d}$ be the algebra $k[X_1, \ldots, X_n]/ I_{d+1}(X_1, \ldots, X_n)$. The multiplication in $P_{n,d}$ can be interpreted as the multiplication of n–variate power series where we only compute with the coefficients of the monomials with degree at most d. For the algebras $P_{n,d}$, the methods of the preceding section give a nice lower bound. The below theorem follows at once from Theorem 1 and the fact that the dimension of $\dim P_{n,d}$ is $\binom{n+d}{n}$.

Theorem 2. *For any $n > 0$ and $d > 0$,*

$$C(P_{n,d}) \geq 3 \cdot \binom{n+d}{n} - \binom{n+\lceil d/2 \rceil}{n} - \binom{n+\lfloor d/2 \rfloor}{n}.$$

Remark 1. If we keep $d > 1$ fixed, then $C(P_{n,d}) \geq (3 - o(1)) \dim P_{n,d}$ (as a function in n). This gives a sequence of explicitly given algebras such that the multiplicative complexity comes arbitrarily close to three times the dimension of the algebra. This is the the best we can expect from currently known lower bounds techniques for bilinear problems. To our knowledge, this is the first lower bound of this kind.

5 Lower Bounds: The Hard Case

Throughout the remainder of this section, we use the following notations: as usual, A is an associative algebra. We denote its multiplication map by ϕ. We assume that we have a decomposition $A = I \oplus X \oplus Y$ (as vector spaces) with vector spaces X and Y and a twosided ideal I. Furthermore, $I^2 = \{0\}$ and $Y \cdot I = \{0\}$. Moreover, let $U \subseteq X$ and $V \subseteq Y$ be vector spaces such that for all projections π of A onto $I \oplus U \oplus V$, the bilinear map $\pi \circ \phi$ is *1–concise*, that is, its left kernel $\{a \in A \mid \pi \circ \phi(a, A) = \{0\}\}$ equals $\{0\}$.

Our general plan looks as follows: using Lemma 2, we reduce the proof of a lower bound for $C(\phi)$ to the proof of a lower bound for $C(\phi|_{X\times I})$ where $\phi|_{X\times I}$ denotes the restriction of ϕ to $X \times I$. This reduction works for any algebra with the above decomposition property. After that, we have to estimate $C(\phi|_{X\times I})$ individually. This is done in the next section for the particularly important case of upper triangular matrices. The main result of this section is the following lower bound.

Lemma 7. *With the notations from above,*

$$C(A) \geq C(\phi|_{X\times I}) + \dim A + \dim Y - \dim U - \dim V.$$

Proof. Let $\beta = (f_1, g_1, w_1, \ldots, f_\ell, g_\ell, w_\ell)$ be a quadratic computation for A. Since $w_1, \ldots, w_\ell$ generate A, we may assume that $\mathrm{lin}\{w_{\ell-m+1}, \ldots, w_\ell\} \oplus I \oplus U \oplus V = A$ where $m = \dim A - \dim I - \dim U - \dim V$. Let π denote the projection along $\mathrm{lin}\{w_{\ell-m+1}, \ldots, w_\ell\}$ onto $I \oplus U \oplus V$. Then $\beta' = (f_1, g_1, w'_1, \ldots, f_{\ell'}, g_{\ell'}, w'_{\ell'})$ with $w'_\lambda = \pi(w_\lambda)$ and $\ell' = \ell - m$ is a quadratic computation for $\pi \circ \phi$.

We claim that β' separates $(A, \{0\}, \{0\})$. If this was not the case, then there would be an $a \in A \setminus \{0\}$ by Lemma 3 such that

$$\pi \circ \phi(a, A) \subseteq \phi(\{0\}, \{0\}) + \{0\} = \{0\}$$

contradicting the assumption that $\pi \circ \phi$ is 1–concise.

From the definition of "separate", it follows that β' also separates $(I \oplus Y, \{0\}, \{0\})$. In other words, $f_1|_{(I\oplus Y)\times\{0\}}, \ldots, f_{\ell'}|_{(I\oplus Y)\times\{0\}}$ generate $((I \oplus Y) \times \{0\})^*$ after possibly exchanging some of the f_λ with the corresponding g_λ.

Let $\psi = (\pi \circ \phi)|_{A\times I}$. Obviously $\hat{\beta} = (\hat{f}_1, \hat{g}_1, w'_1, \ldots, \hat{f}_{\ell'}, \hat{g}_{\ell'}, w'_{\ell'})$ with $\hat{f}_\lambda = f_\lambda|_{A\times I}$ and $\hat{g}_\lambda = g_\lambda|_{A\times I}$ is a quadratic computation for ψ.

As $(I \oplus Y) \times \{0\} \subseteq A \times I$, $\hat{f}_\lambda|_{(I\oplus Y)\times\{0\}} = (f_\lambda|_{A\times I})|_{(I\oplus Y)\times\{0\}} = f_\lambda|_{(I\oplus Y)\times\{0\}}$. From this, we get that also $\hat{\beta}$ separates $(I \oplus Y, \{0\}, \{0\})$. Lemma 2 now yields the lower bound

$$\ell' \geq C(\psi/(I \oplus Y) \times \{0\}) + \dim I + \dim Y. \tag{6}$$

By the definition of "quotient" in Section 2.2, $\psi/(I \oplus Y) \times \{0\}$ is a bilinear map $A/(I \oplus Y) \times I \to (I \oplus U \oplus V)/\tilde{W}$ that maps $(a + (I \oplus Y), b)$ to $\pi \circ \phi(a, b) + \tilde{W}$ where $\tilde{W} = \mathrm{lin}\{\pi \circ \phi(A, \{0\})\} + \mathrm{lin}\{\pi \circ \phi(I \oplus Y, I)\}$. (For a vector space Z, we identify $Z/\{0\}$ with Z.) Since $I^2 = \{0\}$ and $Y \cdot I = \{0\}$ by assumption, $\tilde{W} = \{0\}$. For $x + I \oplus Y \in A/(I \oplus Y)$ and $t \in I$,

$$\psi/(I \oplus Y) \times \{0\}(x + I \oplus Y, t) = \pi \circ \phi(x, t) = xt,$$

since $x \cdot t \in I$. Thus, the following diagram commutes

$$\begin{array}{ccc} A/(I \oplus Y) \times I & \xrightarrow{\psi/(I \oplus Y) \times \{0\}} & I \oplus U \oplus V \\ {\scriptstyle h \times id}\downarrow & & \uparrow \\ X \times I & \xrightarrow{\phi} & I \end{array}$$

where $h : A/(I \oplus Y) \to X$ denotes the canonical isomorphism. Hence we obtain $C(\psi/(I \oplus Y) \times \{0\}) \geq C(\phi|_{X \times I})$.

Exploiting $\ell' = \ell - m$ and choosing β to be an optimal computation, the claim of the lemma follows from (6). $\square$

6 Multiplication of Upper Triangular Matrices

We now apply the results of the preceding section to the algebra $\mathrm{T}_n(k)$ of upper triangular $n \times n$–matrices with entries from k. For the sake of simplicity, we assume that n is even.

In the following, let $e_{i,j} \in \mathrm{T}_n(k)$ denote the matrix that has a one in position (i, j) and zeros elsewhere for $1 \leq i \leq j \leq n$. The radical R of $\mathrm{T}_n(k)$ equals the linear span of all $e_{i,j}$ with $i < j$, that is, R is the set of all matrices with purely zeros on the diagonal. More general, the power R^h equals the linear span of all $e_{i,j}$ with $i + h \leq j$.

We will first step down from $\mathrm{T}_n(k)$ to the quotient $A = \mathrm{T}_n(k)/R^{n/2+1}$. By Corollary 1, we obtain

$$C(\mathrm{T}_n(k)) \geq C(A) + 2 \dim R^{n/2+1} = C(A) + \tfrac{1}{4}n^2 - \tfrac{1}{2}n. \tag{7}$$

The multiplication in A corresponds to the multiplication of upper triangular matrices where we do not compute the entries in the positions (i, j) with $i + n/2 + 1 \leq j$. We use this representation in the following.

Next, we have to instantiate I, X, Y, U, and V. For the remainder of this section, let $m = n/2$. We choose (see Figure 1 for an illustration)

$$\begin{aligned} I &= \mathrm{lin}\{e_{i,j} \mid i \leq m \text{ and } j > m\}, & X &= \mathrm{lin}\{e_{i,j} \mid i \leq m \text{ and } j \leq m\}, \\ Y &= \mathrm{lin}\{e_{i,j} \mid i > m \text{ and } j > m\}, & U &= \{0\}, \\ V &= \mathrm{lin}\{e_{m+1,n}, e_{m+2,n}, \ldots, e_{n,n}\}. \end{aligned}$$

Obviously, $A = I \oplus X \oplus Y$. A straightforward calculation shows that $I^2 = \{0\}$ and $Y \cdot I = \{0\}$ (in A). Moreover, to fulfil the assumptions of Lemma 7, we have to show that for any projection π onto $I \oplus U \oplus V = I \oplus V$, $\pi \circ \phi$ is 1–concise, where ϕ denotes the multiplication map of A. So for each $e_{i,j} \in A$, we have to find an element $a \in A$ such that $\pi(e_{i,j} \cdot a) \neq 0$. We consider three cases: if $e_{i,j} \in I$, that is, $i \leq m$ and $j > m$, then $e_{i,j} \cdot e_{j,j} = e_{i,j} \in I$, thus $\pi(e_{i,j}) = e_{i,j} \neq 0$. If $e_{i,j} \in X$, i.e., $i \leq m$ and $j \leq m$, then $e_{i,j} \cdot e_{j,m+1} = e_{i,m+1} \in I$. If finally $e_{i,j} \in Y$, that is, $i > m$ and $j > m$, then $e_{i,j} \cdot e_{j,n} = e_{j,n} \in V$, thus $\pi(e_{i,j} \cdot e_{j,n}) = e_{j,n} \neq 0$. The 1-conciseness of $\pi \circ \phi$ follows from this.

It remains to estimate $C(\phi|_{X \times I})$. Our aim is to use the following lemma which is proven in [6, Lemma 4]. In what follows, $[B, C] := BC - CB$ denotes the *Lie product* of two matrices.

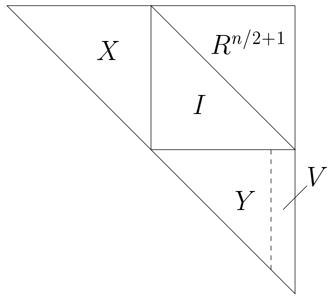

Fig. 1. The decomposition of $\mathrm{T}_n(k)$

Lemma 8. *Let k be a field. Let t be a tensor with 3-slices $I_N, B, C \in k^{N\times N}$. Then $C(t) \geq N + \frac{1}{2}\,\mathrm{rk}[B, C]$.*

To utilize this lemma, we have to determine the tensor of $\phi|_{X\times I}$: for a clearer presentation, we choose the basis

$$\underbrace{e_{1,1},\dots,e_{1,m}}_{\text{first group}},\ \dots,\ \underbrace{e_{i,i},\dots,e_{i,m}}_{i\text{th group}},\ \dots,\ \underbrace{e_{m,m}}_{m\text{th group}}$$

for X (in this row-wise order) and the basis

$$\underbrace{e_{1,m+1},\dots,e_{m,m+1}}_{\text{first group}},\ \dots,\ \underbrace{e_{j,m+j},\dots,e_{m,m+j}}_{j\text{th group}},\ \dots,\ \underbrace{e_{m,n}}_{m\text{th group}}$$

for I (column-wise ordering). The third basis (again for I, since $\Phi|_{X\times I}$ is a mapping $X \times I \to I$) equals the second basis (but we forget about the groups). We denote the 3–slice of the tensor of $\phi|_{X\times I}$ that corresponds to $e_{i,m+j}$ by $T_{i,j}$ for $j \leq i \leq m$. The $T_{i,j}$ are matrices of size $M\times M$ where $M = \frac{1}{2}m(m+1)$. We associate a block structure with the $T_{i,j}$'s induced by the above groups of the first and second basis.

In $T_{i,j}$, the only positions with nonzero entries are in the block at position (i, j), that is, in the positions whose rows and columns correspond to the vectors of the ith and jth group of the above two bases, respectively.

An easy calculation shows that the entries of $T_{i,j}$ within these positions equal

$$\begin{pmatrix} Z_{i-j,j} \\ I_j \end{pmatrix} \tag{8}$$

where I_κ denotes the $\kappa \times \kappa$–identity matrix and $Z_{\mu,\nu}$ denotes the zero matrix of size $\mu \times \nu$. In particular, the 3–slices are linearly independent. From a macroscopic point of view with respect to the above block structure, the $T_{i,j}$ are block lower triangular matrices "of the form $e_{i,j}$" with the above matrix (8) as the only nonzero entry (instead of a one).

By Lemma 1, the fact that $C(\phi|_{X\times I}) \leq \ell$ is equivalent to the existence of rank one matrices $P_1, \ldots, P_\ell$ such that

$$\begin{pmatrix} 0 & T_{i,j} \\ T_{i,j}^\top & 0 \end{pmatrix} \in \operatorname{lin}\{P_1 + P_1^\top, \ldots, P_\ell + P_\ell^\top\} \qquad \text{for } i \leq j \leq m.$$

We now exploit the Steinitz exchange to save one product for each tensor $T_{\mu,\nu}$ with $\mu \geq \nu + 2$: there are matrices $S_1, \ldots, S_m$ and $Q_1, \ldots, Q_{m-1}$ in $\operatorname{lin}\{T_{\mu,\nu} \mid \mu \geq \nu + 2\}$ such that after a suitable permutation of the $P_1, \ldots, P_\ell$

$$\begin{pmatrix} 0 & T_{\mu,\mu} \\ T_{\mu,\mu}^\top & 0 \end{pmatrix} - \begin{pmatrix} 0 & S_\mu \\ S_\mu^\top & 0 \end{pmatrix}, \begin{pmatrix} 0 & T_{\nu,\nu-1} \\ T_{\nu,\nu-1}^\top & 0 \end{pmatrix} - \begin{pmatrix} 0 & Q_\nu \\ Q_\nu^\top & 0 \end{pmatrix} \in$$
$$\operatorname{lin}\{P_1 + P_1^\top, \ldots, P_{\ell-s} + P_{\ell-s}^\top\} \quad \text{for } 1 \leq \mu \leq m, 1 \leq \nu \leq m-1, \tag{9}$$

where $s = \frac{1}{2}m(m+1) - m - (m-1)$. Thus we have killed s products.

Let $\lambda_1, \ldots, \lambda_m \in k$ be pairwise distinct. Define

$$\begin{aligned} E &= T_{1,1} - S_1 + \cdots + T_{m,m} - S_m, \\ B &= \lambda_1(T_{1,1} - S_1) + \cdots + \lambda_m(T_{m,m} - S_m), \\ C &= T_{2,1} - Q_1 + \cdots + T_{m,m-1} - Q_{m-1}. \end{aligned}$$

From (9), we obtain

$$C(\phi|_{X\times I}) \geq s + C(E, B, C). \tag{10}$$

With respect to the above block structure, E has solely identity matrices on the main diagonal and zero matrices one the first subdiagonal. The matrix B has λ_μ multiples of identity matrices on the main diagonal and also zero matrices on the first subdiagonal. The matrix C has zero matrices on the diagonal and "nearly" identity matrices, more precisely, a line of zeros with an identity matrix (as depicted in (8)) one the first subdiagonal.

The matrix E is invertible. By (4),

$$C(E, B, C) = C(I_M, BE^{-1}, CE^{-1}). \tag{11}$$

Due to the structure of E, E^{-1} also has solely identity matrices on the main diagonal and zero matrices on the first subdiagonal. Thus, BE^{-1} has λ_μ multiples of identity matrices on the main diagonal and zero matrices on the first subdiagonal. In the same way, CE^{-1} has zero matrices on the diagonal and "nearly" identity matrices on the first subdiagonal. Some easy algebra shows that due to this structure, the Lie product $[BE^{-1}, CE^{-1}]$ has zero matrices in the blocks on the main diagonal and the matrix

$$\underbrace{(\lambda_{j+1} - \lambda_j)}_{\neq 0} \cdot \begin{pmatrix} Z_{1,j} \\ I_j \end{pmatrix}$$

in the $(j+1, j)$–block (on the first subdiagonal) for $1 \leq j \leq m-1$. Hence, the rank of $[BE^{-1}, CE^{-1}]$ is at least $1 + 2 + \cdots + m - 1 = \frac{1}{2}(m-1)m$.

Together with (10), (11), and Lemma 8, the last statement implies the following lower bound.

Lemma 9. *With the notations from above,* $C(\phi|_{X\times I}) \geq \frac{5}{4}m^2 - \frac{5}{4}m + 1$.

Exploiting (7) and then bounding $C(A)$ by Lemma 7 and Lemma 9, we obtain the following lower bound.

Theorem 3. *For even* n*, the multiplicative complexity of the multiplication of upper triangular matrices of size* $n \times n$ *has the lower bound*

$$C(\mathrm{T}_n(k)) \geq \tfrac{17}{16}n^2 - \tfrac{3}{8}n + 1 \geq (2\tfrac{1}{8} - o(1)) \dim \mathrm{T}_n(k).$$

Remark 2. For odd n, the same approach also yields $(2\frac{1}{8} - o(1)) \dim \mathrm{T}_n(k)$ as a lower bound. A quick solution goes a follows: simply embed $\mathrm{T}_{n-1}(k)$ into $\mathrm{T}_n(k)$ and apply the above theorem. We only loose an additive amount of $O(n) \leq o(\dim \mathrm{T}_n(k))$.

References

1. A. Alder and V. Strassen. On the algorithmic complexity of associative algebras. *Theoret. Comput. Sci.*, 15:201–211, 1981.
2. Markus Bläser. Bivariate polynomial multiplication. In *Proc. 39th Ann. IEEE Symp. on Foundations of Comput. Sci. (FOCS)*, pages 186–191, 1998.
3. Markus Bläser. Lower bounds for the multiplicative complexity of matrix multiplication. *Comput. Complexity*, 8:203–226, 1999.
4. Markus Bläser. *Untere Schranken für den Rang assoziativer Algebren*. Dissertation, Universität Bonn, 1999.
5. Markus Bläser. Lower bounds for the bilinear complexity of associative algebras. *Comput. Complexity*, 9:73–112, 2000.
6. Markus Bläser. A $\frac{5}{2}n^2$–lower bound for the multiplicative complexity of $n\times n$–matrix multiplication. In *Proc. 18th Int. GI–MIMD Symp. on Theoret. Aspects of Comput. Sci. (STACS)*, Lecture Notes in Comput. Sci. 2010, pages 99–110, 2001.
7. Werner Büchi. *Über eine Klasse von Algebren minimalen Rangs*. Dissertation, Universität Zürich, 1984.
8. Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic Complexity Theory*. Springer, 1997.
9. Yurij A. Drozd and Vladimir V. Kirichenko. *Finite Dimensional Algebras*. Springer, 1994.
10. Hans F. de Groote. *Lectures on the Complexity of Bilinear Problems*. Lecture Notes in Comput. Science 245. Springer, 1986.
11. Joos Heintz and Jacques Morgenstern. On associative algebras of minimal rank. In *Proc. 2nd Applied Algebra and Error Correcting Codes Conf. (AAECC)*, Lecture Notes in Comput. Sci. 228, pages 1–24. Springer, 1986.
12. Joseph Ja'Ja'. On the complexity of bilinear forms with commutativity. *SIAM J. Comput.*, 9:717–738, 1980.
13. Victor Ya. Pan. Methods for computing values of polynomials. *Russ. Math. Surv.*, 21:105–136, 1966.
14. Richard S. Pierce. *Associative Algebras*. Springer, 1982.
15. Volker Strassen. Vermeidung von Divisionen. *J. Reine Angew. Math.*, 264:184–202, 1973.
16. Volker Strassen. Algebraic complexity theory. In J. van Leeuven, editor, *Handbook of Theoretical Computer Science Vol. A*, pages 634–672. Elsevier Science Publishers B.V., 1990.