On the Multilinear Complexity of Associative Algebras

Markus Bläser

Universität des Saarlandes, Saarland Informatics Campus, Saarbrücken, Germany

Hendrik Mayer[1]

Massachusetts Institute of Technology, Cambridge, MA, USA

Devansh Shringi[1]

University of Toronto, Canada

Abstract

Christandl and Zuiddam [12] study the multilinear complexity of d-fold matrix multiplication in the context of quantum communication complexity. Bshouty [8] investigates the multilinear complexity of d-fold multiplication in commutative algebras to understand the size of so-called testers. The study of bilinear complexity is a classical topic in algebraic complexity theory, starting with the work by Strassen. However, there has been no systematic study of the multilinear complexity of multilinear maps.

In the present work, we systematically investigate the multilinear complexity of d-fold multiplication in arbitrary associative algebras. We prove a multilinear generalization of the famous Alder-Strassen theorem, which is a lower bound for the bilinear complexity of the (2-fold) multiplication in an associative algebra. We show that the multilinear complexity of the d-fold multiplication has a lower bound of $d \cdot \dim A - (d-1)t$, where t is the number of maximal twosided ideals in A. This is optimal in the sense that there are algebras for which this lower bound is tight. Furthermore, we prove the following dichotomy that the quotient algebra $A/\operatorname{rad} A$ determines the complexity of the d-fold multiplication in A: When the semisimple algebra $A/\operatorname{rad} A$ is commutative, then the multilinear complexity of the d-fold multiplication in A is polynomial in d. On the other hand, when $A/\operatorname{rad} A$ is noncommutative, then the multilinear complexity of the d-fold multiplication in A is exponential in d.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory; Mathematics of computing

Keywords and phrases Multilinear computations, associative algebras, matrix multiplication, Alder–Strassen theorem

Digital Object Identifier 10.4230/LIPIcs.STACS.2023.12

Acknowledgements Hendrik Mayer would like to thank the MIT MISTI-Germany program for funding his internship.

1 Introduction

A fundamental problem in algebraic complexity theory is the question about the costs of multiplication, for instance, of matrices, triangular matrices, or polynomials. To be more specific, let $\mathbb{F}$ be a field and let A be a finite dimensional associative $\mathbb{F}$-algebra with unity 1. By fixing a basis of A, say $v_1, \dots, v_N$, 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 $1 \le \mu, \nu \le N$ with *structural constants* $\alpha_{\mu,\nu,\kappa} \in \mathbb{F}$, then these constants and the identity

[1] Work done during an internship at Saarland University.

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

define the desired bilinear forms $b_1, \ldots, b_N$. The *bilinear complexity* or *rank* of $b_1, \ldots, b_N$ is the smallest number of *bilinear* multiplications necessary and sufficient to compute $b_1, \ldots, b_N$ from the indeterminates $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_N$. A bilinear multiplication is a multiplication of the form $u_\rho(X_1, \ldots, X_N)v_\rho(Y_1, \ldots, Y_N)$, where u_ρ and v_ρ are linear forms. Additions and multiplications with scalars from $\mathbb{F}$ are free of costs.

It is easy to see that the bilinear 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 bilinear complexity (or rank) of (the multiplication in) A. Equivalently, we can formulate the problem as a tensor rank problem. Given the structure tensor $(\alpha_{\mu,\nu,\kappa})$ of the algebra, which you can think of as a three-dimensional matrix, the bilinear complexity of an algebra is exactly the number of rank-one tensors that are needed to write the structure tensor of an algebra as a sum of rank-one tensors. For an introduction to algebraic complexity theory and for further background on tensor rank, we recommend [11, 18].

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

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

where t is the number of maximal twosided ideals in A. (See Section 2 for more background on algebras.) This bound has been improved for a large class of so-called semisimple algebras to $\frac{5}{2}\dim A - 3n$, where n is the sum of the sizes of the matrices in the decomposition of A into simple algebras [3]. The Alder–Strassen theorem itself even holds for a more general complexity measure, the so-called multiplicative complexity. Algebras for which the Alder–Strassen bound is tight are called algbras of minimal rank or minimal multiplicative complexity. They have been characterised in terms of their algebraic structure in [4] and [6].

The most prominent algebra is the algebra of $n \times n$-matrices. Strassen [23] proved that the rank of 2×2-matrix multiplication is upper bounded by 7, giving rise to his famous matrix multiplication algorithm. Winograd [25] proved that the bound of 7 is optimal, which can be considered as a first instance of the Alder–Strassen theorem. Over the past decades, an exciting development of fast matrix multiplication algorithms has taken place, culminating in the current fastest algorithms with running time $O(n^{2.373})$ [14, 22, 24, 16, 2], see [5] for an overview. The best lower bound is due to Landsberg [17] and is $R(\mathbb{F}^{n\times n}) \geq 3n^2 - o(n^2)$.

When an algebra A is associative, the d-fold multiplication is a multilinear map $A \times \cdots \times A \to A$. The notion of bilinear complexity naturally generalizes: Instead of bilinear products, we have d-linear products. Equivalently, we can study the tensor rank of tensors of higher orders. Christandl and Zuiddam [12] recently studied the multilinear complexity of d-fold matrix multiplication, which is a so-called graph tensor on the cycle graph of length d. The multilinear complexity of such graph tensors plays a particular role in quantum communication complexity, see e.g. [10]. Prior to this, Nisan and Wigderson studied depth-three circuits for iterated matrix multiplication using the partial derivative method [19].

Bshouty [9] invented the concept of testers. Testers are a useful tool in algebraic algorithms to reduce identity testing from large domains to smaller ones. In the case of the class of multilinear forms of degree d over an algebra A, he proves that the size of the optimal tester is exactly equal to the multilinear complexity of the d-fold multiplication in the algebra A [8]. Thus, it is interesting to understand the multilinear complexity of d-fold multiplication of arbitrary algebras.

Related models have been studied in algebraic complexity theory, like (set-multilinear) depth-3-circuits, higher order tensor rank, or Waring rank, see e.g. the surveys [21, 7]. But there has been no systematic study of the multilinear complexity of d-fold multiplication maps.

1.1 Our work

We initiate the systematic study of the multilinear complexity of d-fold multiplication in an associative algebra. Our motivation comes from the work mentioned above that relies on bounds for the multilinear complexity in certain algebras.

We prove a multilinear generalization of the Alder–Strassen theorem (Theoreom 19), namely that

$$R(A, d) \geq d \cdot \dim A - (d-1)t,$$

where t is the number of maximal twosided ideals in A. Here $R(A, d)$ denotes the rank of d-fold multiplication in A (see Section 3). This bound is tight in the sense that there are algebras for which equality holds, for instance, products of simply generated division algebras (see Section 5.3). For $d = 2$, we exactly recover the Alder–Strassen theorem.

Moreover, an interesting phenomenon arises. When we keep the algebra A fixed and consider the multilinear complexity $R(A, d)$ as a function of d, then the growth is either polynomial (where the exponent might depend on A) or exponential. When $A/\operatorname{rad} A$ is commutative, then $R(A, d)$ is polynomial, more precisely $R(A, d) \leq d^{(s-1)(D+1)} \cdot D^{(D+3)s}$, where D is the dimension of A and s is the so-called index of nilpotency, which can be upper bounded by D (Theorem 18). This result holds over large enough perfect fields. Note that most fields are perfect, like fields of characteristic zero, finite fields, or algebraically closed fields. On the other hand, when $A/\operatorname{rad} A$ is noncommutative, then we prove an exponential lower bound (Theorem 13 together with Lemma 5). So we obtain a dichotomy.

We would like to stress once again that the motivation for this work are the above mentioned applications in quantum communication and the study of testers. Multilinear computations are in general not suited to actually evaluate d-fold multiplication maps in practice. While for instance Christandl and Zuiddam [12] prove that the multilinear complexity of d-fold matrix multiplication is exponential in d, we can of course multiply d matrices in polynomial time. The reason is that in practice, we first multiply two matrices, then multiply the result with the third one and so on. In multilinear computations, we cannot reuse results. It is essentially the same as the difference between formula and circuit size, which can be exponential, too. Nevertheless, bilinear algorithms have been successfully used to construct algorithms for the multiplication of two matrices, because in the case $d = 2$, bilinear computations and circuit size only differ by a factor of 2, see e.g. [5].

1.2 Organisation of the paper

Section 2 provides some basic facts about associative algebras. In Section 3, we introduce the model of computation and prove some basic facts about it. Then we prove how to remove the radical in Section 4, similar to the original proof of the Alder–Strassen theorem. In Sections 5 and 6, we study semisimple algebras, first semisimple algebras that are in addition basic and then arbitrary semisimple algebras. Then we go on with a study of the algebra of upper triangular matrices in Section 7, which is an example of a noncommutative algebra such that $A/\operatorname{rad} A$ is commutative. In constrast to general matrices, for which we have an exponential lower bound, we prove a polynomial upper bound for the multilinear complexity (as a function of d). In Section 8, we generalize this result to arbitrary, potentially noncommutative algebras A for which $A/\operatorname{rad} A$ is commutative. Finally, we prove the generalized Alder–Strassen theorem in Section 9.

2 Structure of associative algebras

We collect some elementary properties of associative algebras. The term *algebra* here always means a finite dimensional associative algebra with identity 1 over some field $\mathbb{F}$. The term *left module* and *right module* always means a finitely generated left module and right module over some algebra A, respectively. By the embedding $\alpha \mapsto \alpha \cdot 1$, $\mathbb{F}$ becomes a subalgebra of A. Hence, every A-left module or A-right module is also a finite dimensional $\mathbb{F}$-vector space. If we speak of a basis of an algebra or a module, we always mean a basis of the underlying vector space. Further material as well as proofs of the mentioned properties can be found in [20, 13, 15].

A left ideal I (and in the same way, a right ideal or twosided ideal) is called *nilpotent*, if $I^n = \{0\}$ for some positive integer n.

▶ **Fact 1.** *For all finite dimensional algebras A, the following holds:*

1. *The sum of all nilpotent left ideals of A is a nilpotent twosided ideal, which contains every nilpotent right ideal of A. This twosided ideal is called the* radical *of A and is denoted by* $\operatorname{rad} A$.
2. *The quotient algebra $A/\operatorname{rad} A$ contains no nilpotent ideals other than the zero ideal.*
3. *The radical of A is contained in every maximal twosided ideal of A.*
4. *The algebras A and $A/\operatorname{rad} A$ have the same number of maximal twosided ideals.*

We call an algebra A *semisimple*, if $\operatorname{rad} A = \{0\}$. By the above fact, $A/\operatorname{rad} A$ is semisimple. An algebra A is called *simple*, if there are no twosided ideals in A, except the zero ideal and A itself.

We now describe some of the most important ways to construct new algebras from given ones: If A and B are $\mathbb{F}$-algebras, then the direct product $A \times B$ with componentwise addition and multiplication is again an $\mathbb{F}$-algebra. The set of all $n \times n$–matrices with entries from A forms an $\mathbb{F}$-algebra (with the usual definition of addition and multiplication of matrices). This algebra is denoted by $M_n(A)$ or $A^{n\times n}$.

We denote the set of all units of an algebra A, that is, the set of all invertible elements, by $A^\times$. An algebra D is called a *division algebra*, if $D^\times = D \setminus \{0\}$. An algebra A is called *local*, if $A/\operatorname{rad} A$ is a division algebra, and A is called *basic*, if $A/\operatorname{rad} A$ is a direct product of division algebras.

If $x \in A$, we denote by AxA the ideal generated by x. If A is commutative, we will also write (x) for short. Furthermore, $\mathbb{F}[x]$ denotes the smallest subalgebra of A that contains x. If $x_1, \dots, x_m \in A$ mutually commute, then $\mathbb{F}[x_1, \dots, x_m]$ denotes the smallest subalgebra of A that contains $x_1, \dots, x_m$. For elements $v_1, \dots, v_n$ of some vector space, $\langle v_1, \dots, v_n \rangle$ denotes their linear span.

The following fundamental theorem describes the structure of semisimple algebras.

▶ **Theorem 2** (Wedderburn). *Every finite dimensional semisimple algebra is isomorphic to a finite direct product of simple algebras. Every finite dimensional simple $\mathbb{F}$-algebra A is isomorphic to an algebra $M_n(D)$ for an integer $n \geq 1$ and an $\mathbb{F}$-division algebra D. The integer n and the algebra D are uniquely determined by A (the latter one up to isomorphism).*

3 Multilinear computations

For a vector space V, V^* denotes the dual space of V, that is, the vector space of all linear forms on V.

We define multilinear complexity in a coordinate-free way. The d-fold multiplication in an algebra is a multilinear map $\phi : A^d \to A$. A multilinear computation consists of linear forms $F_{i,\delta} \in A^*$, $1 \leq \delta \leq d$, and elements $w_i \in A$, $1 \leq i \leq r$, such that

$$x_1 \cdot x_2 \cdot \ldots \cdot x_d = \sum_{i=1}^{r} F_{i,1}(x_1) F_{i,2}(x_2) \ldots F_{i,d}(x_d) \cdot w_i$$

for all $x_1, \ldots, x_d \in A$. r is called the length of the computation. The length of a shortest computation for the d-fold multiplication is the d-linear complexity of A. We denote this quantity by $R(A, d)$.

In the remainder of this section, we collect some useful properties of multilinear complexity. If we choose vector spaces $V_i \subseteq A$, $1 \leq \delta \leq d$, we get a multilinear map $\phi' : V_1 \times \ldots, \times V_d \to A$ in the canonical way. By restricting each $F_{i,\delta}$ to V_δ, the computation above turns into a computation for ϕ' of the same length. Obviously, $R(\phi') \leq R(\phi)$.

If I is a twosided ideal of A, then A/I is an algebra again. Each $F_{i,\delta}$ induces a linear form on A/I in the canonical way. By replacing each $F_{i,\delta}$ with this linear form and mapping each w_i to its image in A/I, the computation turns into a computation for the multiplication in A/I.

We can define an equivalence relation on the set of all d-linear computations for an algebra. Let $a_0, \ldots, a_d \in A$ be invertible. We have the identity

$$a_0^{-1} a_0 x_1 a_1^{-1} a_1 x_2 a_2^{-1} a_2 \ldots a_{d-1} x_d a_d{}^{-1} a_d = x_1 \ldots x_d.$$

Therefore, the computation given by $\hat{F}_{i,\delta}(x) = F_{i,\delta}(a_{i-1} x a_i^{-1})$, $1 \leq \delta \leq d$, and $\hat{w}_i = a_0^{-1} w_i a_d$ is again a computation for ϕ, the multiplication in A. The action of $(A^\times)^{d+1}$ defines an equivalence relation on the set of all computations of length r.

The following claim will turn out useful in our lower bound proofs:

▷ Claim 3. Consider a computation for an algebra A of dimension N as above. For every j, $F_{1,j}, F_{2,j}, \ldots, F_{r,j}$ span A^*. That is, we can have $F_{1,j}, \ldots, F_{N,j}$ as a basis of A^* after reordering.

Proof. Assume they do not span A^* for some j, then there is an element $y \in A \setminus \{0\}$ such that $F_{1,j}(y) = F_{2,j}(y) = \ldots = F_{r,j}(y) = 0$. This means that $x_1 \cdot x_2 \cdot \ldots \cdot x_{j-1} \cdot y \cdot x_{j+1} \cdot \ldots \cdot x_d = 0$ for all $x_1, \ldots, x_d$. By setting $x_i = 1$ for $i \neq j$, we get that $y = 0$, which is a contradiction. We can reorder the $F_{i,j}$'s such that $F_{1,j}, \ldots, F_{N,j}$ span A^*. ◁

The first item of the following fact follows from the trivial decomposition. The second by setting $x_{d+1} = 1$.

▶ **Fact 4.**
1. $R(A, d) \leq (\dim A)^d$
2. $R(A, d) \leq R(A, d + 1)$

The d-fold multiplication in an algebra corresponds to a tensor in $t_{A,d} \in A^* \otimes \cdots \otimes A^* \otimes A$. The rank of $t_{A,d}$ is the minimum number r of rank-one tensors $u_{\rho,1} \otimes \cdots \otimes u_{\rho,d} \otimes v_\rho \in A^* \otimes \cdots \otimes A^* \otimes A$ such that

$$t_{A,d} = \sum_{\rho=1}^{r} u_{\rho,1} \otimes \cdots \otimes u_{\rho,d} \otimes v_\rho.$$

Tensor rank and multilinear complexity coincide, that is, $R(t_{A,d}) = R(A, d)$.

4 Removing the radical $\operatorname{rad}(A)$

We start by generalizing the first lemma of the proof by Alder-Strassen to multilinear complexity, allowing us to work with the semisimple algebra $A/\operatorname{rad}(A)$, which has a nicer structure than a general algebra.

▶ **Lemma 5.**

$$R(A,d) \geq R(A/\operatorname{rad}(A),d) + d \cdot \dim(\operatorname{rad}(A)).$$

Proof. Consider an algebra A over the field $\mathbb{F}$. Let ϕ be a length r computation with $F_{i,j} \in A^*$ and $w_i \in A$ such that the d-fold multiplication of A is computed by the following equation:

$$x_1 \cdot x_2 \cdot \ldots \cdot x_d = \sum_{i=1}^{r} F_{i,1}(x_1)F_{i,2}(x_2)\ldots F_{i,d}(x_d) \cdot w_i.$$

We will inductively construct vector spaces $V_1, \ldots, V_d$ such that $V_\delta \oplus \operatorname{rad}(A) = A$ and $1 \in V_\delta$. Let ϕ_j be the multiplication map restricted to $V_1 \times \cdots \times V_j \times A \times \cdots \times A$. We will now prove that $R(A,d) \geq R(\phi_j) + j \cdot \dim(\operatorname{rad}(A))$. The base case $j = 0$ is clear.

Induction Hypothesis: $R(A,d) \geq R(\phi_{j-1}) + (j-1) \cdot \dim(\operatorname{rad}(A))$.

Induction Step: We obtain a basis of A^* using the following claim similar to Claim 3, which we prove later.

▷ Claim 6. For all j, we have that in a computation for ϕ_{j-1}, $F_{1,j}, F_{2,j}, \ldots, F_{r,j}$ span A^*. That is, we can have $F_{1,j}, \ldots, F_{N,j}$ as a basis of A^* after reordering. Here, $N = \dim A$.

Now for the basis $F_{1,j}, \ldots, F_{N,j}$ of A^*, we calculate the dual basis $a_{1,j}, \ldots, a_{N,j}$, which is a basis of A as $A^{**} = A$ (A is finite dimensional). Note that from the definition of dual basis, $F_{i,j}(a_{k,j}) = \delta_{ik}$, where δ_{ik} is Kronecker's delta.

Consider the canonical projection $P : A \to A/\operatorname{rad}(A)$. Let $\dim(\operatorname{rad}(A)) = \rho$. Then $\dim(A/\operatorname{rad}(A)) = N - \rho$. After rearrangement, we can have that $P(a_{1,j}), \ldots, P(a_{N-\rho,j})$ form a basis of $A/\operatorname{rad}(A)$. So we can assume w.l.o.g. that $P(a_{i,j})$ for $i \in [N-\rho]$ span $A/\operatorname{rad}(A)$. In particular, with $V_j := \langle a_{1,j}, \ldots, a_{N-\rho,j} \rangle$, we have that $V_j \cap \operatorname{rad} A = \{0\}$.

We also observe that $1 \notin \operatorname{rad}(A)$. Now, we want that $a_{1,j}, \ldots, a_{N-\rho,j}$ span 1. If they don't, then there must exist $z \in \operatorname{rad}(A)$ such that $1 - z$ is spanned by $a_{1,j}, \ldots, a_{N-\rho,j}$, i.e., there are $\beta_1, \ldots, \beta_{N-\rho} \in \mathbb{F}$ such that

$$1 - z = \sum_{i=1}^{N-\rho} \beta_i a_{i,j}.$$

As $z \in \operatorname{rad}(A)$, there exists an s such that $z^s = 0$. Therefore, $1 - z$ is invertible and has inverse $1 + z + z^2 + \ldots + z^{s-1}$. We can write

$$x_1 \ldots x_{j-1} x_j x_{j+1} \ldots x_d = x_1 \ldots x_{j-1}(x_j(1-z))((1-z)^{-1}x_{j+1}) \ldots x_d,$$

which changes the computation as follows:

$$\forall i \in [r] : \ \hat{F}_{i,j}(x_j) = F_{i,j}(x_j(1-z)), \text{ and } \hat{F}_{i,j+1}(x_{j+1}) = F_{i,j+1}((1-z)^{-1}x_{j+1}).$$

All other $F_{i,h}$ are not changed. If $j = d$, then w_i is changed instead of $F_{i,d+1}$ (which does not exist). Compare also Section 3.

The dual basis will change as $\hat{a}_{i,j} = a_{i,j}(1-z)^{-1}$. As we have $1-z$ in the span of $a_{1,j}, \ldots, a_{N-\rho,j}$, we will have that 1 is in the span of $\hat{a}_{1,j}, \ldots, \hat{a}_{N-\rho,j}$. Note that this does not change the spaces $V_1, \ldots, V_{j-1}$. So, we have $1 = \sum_{i=1}^{N-\rho} \beta_i \hat{a}_{i,j}$.

Set $V_j = \langle \hat{a}_{1,j}, \ldots, \hat{a}_{N-\rho,j} \rangle$. By the choice of V_j, $F_{i,j}|_{V_j} = 0$ for $i \in \{N-\rho+1, \ldots N\}$, as $\hat{F}_{i,j}(\hat{a}_{k,j}) = F_{i,j}(a_{k,j}) = 0$ if $i \neq k$. Thus when restricting to V_j, we can remove ρ products from the computation. By the induction hypothesis, we get that

$$R(A,d) \geq R(\phi_{j-1}) + (j-1)\dim(\operatorname{rad} A)) \geq R(\phi_j) + \dim(\operatorname{rad} A) + (j-1)\dim(\operatorname{rad} A)).$$

This finishes the proof of the induction step.

We now have a multilinear map $\phi_d : V_1 \times \cdots \times V_d \to A$ with $R(A,d) \geq R(\phi_d) + d \cdot \dim(\operatorname{rad} A))$. To finish the proof of the lemma, it suffices to prove $R(\phi_d) \geq R(A/\operatorname{rad} A, d)$. Since $V_\delta \cap \operatorname{rad} A = \{0\}$ for $\delta \in [d]$, the restriction of the projection P to V_δ is an isomorphism. The following diagram commutes:

$$\begin{array}{ccccccc} V_1 & \times & \ldots & \times & V_d & \longrightarrow & A \\ \downarrow & & \ldots & & \downarrow & & \downarrow \\ A/\operatorname{rad} A & \times & \ldots & \times & A/\operatorname{rad} A & \longrightarrow & A/\operatorname{rad} A \end{array}$$

Thus a computation for ϕ_d can be turned into a computation for the d-fold multiplication in $A/\operatorname{rad} A$ (compare Section 3). ◀

We also give a proof of Claim 6, which we made in the proof above.

Proof of Claim 6. Assume they do not span A^* for some j, then there is an element $y \in A\setminus\{0\}$ such that $F_{1,j}(y) = F_{2,j}(y) = \ldots = F_{r,j}(y) = 0$. This means that $x_1 \cdot x_2 \cdot \ldots \cdot x_{j-1} \cdot y \cdot x_{j+1} \cdot \ldots \cdot x_d = 0$ for all $x_1, \ldots, x_d$. Since $1 \in V_\delta$ for every δ, this means $y = 0$, which is a contradiction. We can reorder the $F_{i,j}$'s such that $F_{1,j}, \ldots, F_{N,j}$ span A^*. ◁

4.1 A tight example

An example we can consider is the algebra $A = \mathbb{F}[x]/(x^n)$ of univariate polynomials modulo x^n. We clearly have $\dim(A) = n$. We see that all polynomials in the algebra with zero constant term are nilpotent. Therefore, $\operatorname{rad}(A) = \langle x, x^2, \ldots, x^{n-1} \rangle$ and $\dim(\operatorname{rad}(A)) = n-1$.

From Lemma 5, we get that $R(A,d) \geq R(A/\operatorname{rad}(A), d) + d \cdot \dim(\operatorname{rad}(A))$, i.e. for this case $R(A,d) \geq d \cdot (n-1) + 1$ as $A/\operatorname{rad}(A)$ is just $\mathbb{F}$.

For the upper bound, we see that the d-fold product (without reduction modulo x^n) $f = f_1 \cdot f_2 \cdot \ldots \cdot f_d$ will actually be a polynomial of degree $d(n-1)$. We can evaluate each f_i at $d(n-1)+1$ points (which is free of costs in our model) and multiply the evaluations with $d(n-1)+1$ costing multiplication operations to get an evaluation of f at $d(n-1)+1$ points. (This assumes that our field $\mathbb{F}$ is large enough.) Using $d(n-1)+1$ evaluations of f, we can obtain f using interpolation, which is again free of costs. We ignore the terms with degree higher than n, thus obtaining the final result f. This gives an upper bound of $d \cdot (n-1) + 1$, proving that our lower bound is tight in this case.

5 Products of division algebras

Now that we are to work on $R(A/\operatorname{rad}(A), d)$, we see that $A/\operatorname{rad}(A)$ is a semisimple algebra. Using Theorem 2, we have

$$A/\operatorname{rad}(A) = A_1 \times A_2 \times \ldots \times A_k$$

where $A_i = M_{n_i}(D_i)$, i.e., $n_i \times n_i$ matrices with entries from a division algebra D_i.

In this section, we will mainly focus on case when $n_i = 1$, i.e., all A_i are division algebras.

5.1 Single division algebra ($k = 1$)

For simplicity, we will first take a look at the following lemma, for the case when $k = 1$, i.e., $A = D$ where D is any division algebra. Let $N = \dim(A)$.

▶ **Lemma 7.** *Let A be a division algebra. Then*

$$R(A, d) \geq d \cdot \dim(A) - (d - 1).$$

Proof. Let $r = R(A, d)$. We consider an optimal length r computation for the d-fold multiplication of A. By Claim 3, $F_{1,1}, \dots, F_{N,1}$ is a basis of A^*. Let $a_{1,1}, \dots, a_{N,1}$ be the dual basis, that is, $F_{i,1}(a_{j,1}) = \delta_{i,j}$. We know that the $a_{1,1}, \dots, a_{N,1}$ are all invertible. Let us define a new basis $\hat{a}_{i,1} = a_{i,1}{a_{1,1}}^{-1}$, $i \in [N]$. Let

$$\forall i \in [r] : \hat{F}_{i,1}(x_1) = F_{i,1}(x_1 a_{1,1}), \ \ \hat{F}_{i,2}(x_2) = F_{i,2}({a_{1,1}}^{-1} x_2), \text{ and } \hat{F}_{i,j}(x_j) = F_{i,j}(x_j), \, j \geq 3.$$

We note that $\hat{a}_{1,1} = 1$ and $\hat{F}_{i,1}(1) = F_{i,1}(a_{1,1}) = \delta_{i,1}$. We have that:

$$1 \cdot x_2 \cdot \ldots \cdot x_d = \hat{F}_{1,1}(\hat{a}_{1,1})\hat{F}_{1,2}(x_2) \ldots \hat{F}_{1,d}(x_d) \cdot w_1 + \sum_{i=N+1}^{r} \hat{F}_{i,1}(\hat{a}_{1,1})\hat{F}_{i,2}(x_2) \ldots \hat{F}_{i,d}(x_d) \cdot w_i.$$

As there are $r - N + 1$ terms to sum, $R(A, d - 1) \leq r - N + 1 = R(A, d) - N + 1$. Since $R(A, 1) = N$, we have by induction that

$$R(A, d) \geq N + (N - 1)(d - 1) = d \cdot \dim(A) - (d - 1).$$ ◀

5.2 Arbitrary products of division algebras

Now $A = D_1 \times \cdots \times D_k$ is a product of division algebras.

▶ **Theorem 8.**

$$R(A, d) \geq d \cdot \dim(A) - (d - 1)k.$$

Proof. As in the single division algebra case, we consider an optimal length r computation for the d-fold multiplication in A:

$$x_1 \cdot x_2 \cdot \ldots \cdot x_d = \sum_{i=1}^{r} F_{i,1}(x_1)F_{i,2}(x_2) \ldots F_{i,d}(x_d) \cdot w_i.$$

Similarly to the single division algebra case, $F_{1,1}, F_{2,1}, \dots, F_{N,1}$ is w.l.o.g. a basis of A^*. Let $a_1, a_2, \dots, a_N$ be the dual basis, that is, $F_{i,1}(a_j) = \delta_{i,j}$. Each a_i can be written as $a_i = (a_{i,1}, a_{i,2}, \dots, a_{i,k})$, where each $a_{i,\ell} \in D_\ell$, $1 \leq \ell \leq k$.

We can assume w.l.o.g. that the span of $a_1, \dots, a_k$ contains an element $b = (b_1, \dots, b_k)$ that is nonzero in every component of $D_1 \times \cdots \times D_k$. Since each b_ℓ is nonzero, it is invertible, because D_ℓ is a division algebra. Thus b is invertible, too, and $b^{-1} = (b_1^{-1}, \dots, b_k^{-1})$.

Now, we define $\hat{a}_j = a_j b^{-1}$. Also, let $\hat{F}_{i,1}(x_1) = F_{i,1}(x_1 b)$ and $\hat{F}_{i,2} = F_{i,2}(b^{-1} x_2)$. By construction, $1 = (1, \dots, 1)$ is contained in the linear span of $\hat{a}_1, \dots, \hat{a}_k$. Thus

$$1 \cdot x_2 \cdot \ldots \cdot x_d = \sum_{\ell=1}^{k} \hat{F}_{\ell,1}(1)\hat{F}_{\ell,2}(x_2) \ldots \hat{F}_{\ell,d}(x_d) \cdot w_\ell + \sum_{i=N+1}^{r} \hat{F}_{i,1}(1)\hat{F}_{i,2}(x_2) \ldots \hat{F}_{i,d}(x_d) \cdot w_\ell.$$

Similarly to the $k = 1$ case, we have that $R(A, d - 1) \leq r - N + k = R(A, d) - N + k$. Using induction, we get $R(A, d) \geq N + (d - 1)N + (d - 1)k$, because $R(A, 1) = N$. ◀

5.3 Tight example

An example we can consider is the divison algebra $A = \mathbb{F}[x]/(x^n + 1)$, with $x^n + 1$ being irreducible (think of $\mathbb{F} = \mathbb{Q}$ for instance). We clearly have $\dim(A) = n$. This example is similar to the one in Section 4.1.

From Lemma 7, we get that $R(A, d) \geq d \cdot \dim(A) - (d - 1)$. For the matching upper bound, we see that $f = f_1 \cdot f_2 \cdot \ldots \cdot f_d$ will actually be a polynomial of degree $d(n - 1)$. We can evaluate each f_i at $d(n - 1) + 1$ different points and multiply the evaluations to get an evaluation of f at $d(n - 1) + 1$ points. Using these $d(n - 1) + 1$ evaluations of f, we can obtain f using interpolation. Again, $\mathbb{F}$ needs to be large enough. Finally, we calculate $f \bmod (x^n + 1)$, giving the final result in A. Thus, this yields an upper bound of $d \cdot (n - 1) + 1 = d \cdot \dim A - (d - 1)$.

If we take the k-fold product A^k, we also see that the bound of Theorem 8 is tight, too, for arbitrary k.

In general, we can get a polynomial upper bound (in d) on the multilinear complexity of the d-fold multiplication in any commutative algebra, provided that the field $\mathbb{F}$ is large enough. The exponent of the polynomial depends on the dimension of the algebra.

▶ **Fact 9.** *Let A be a finite dimensional commutative algebra. Then there is a polynomial ring R and an ideal $I \subset R$ such that $A = R/I$.*

▶ **Lemma 10.** *Let A be a commutative algebra over a field $\mathbb{F}$ with k generators and highest degree δ of a variable in a basis of the vector space R/I as above. Then $R(A, d) \leq (d\delta + 1)^k$ provided that $|\mathbb{F}| \geq d\delta + 1$.*

Proof. The idea is to use the same construction as in the example above. Each element x_i in the d-fold product can be presented as a polynomial in k variables and degree at most δ in each variable (modulo I). When we multiply these polynomials, we get a polynomial of individual degree $\leq d\delta$ in k variables. By multivariate Lagrange interpolation, it is enough to interpolate on a k-dimensional grid with $(d\delta + 1)$ points in each direction and a total number of $(d\delta + 1)^k$ many points. In this way we get the stated upper bound because reduction modulo I is free of costs in our model. ◀

We can bound k and $\delta - 1$ in the lemma above by $\dim A$. Thus the bound simplifies to $(d \cdot \dim A)^{\dim A}$.

When A is simply generated, then the construction of the lemma becomes the construction of the example above and we get the following tight bound.

▶ **Corollary 11.** *If A is a simply generated commutative algebra, then $R(A, d) = d \cdot \dim A - (d - 1)$.*

Let D be a commutative division algebra, that is, an extension field. If D is separable, then by the primitive element theorem, it is simply generated and the corollary above applies.

6 General semisimple algebras

In this section, we will look at the case of semisimple algebras, that is $A = A_1 \times \cdots \times A_k$ and each A_ℓ is a matrix algebra of format $n_\ell \times n_\ell$ with entries from a divison algebra D_ℓ. We can assume that at least one $n_\ell > 1$, otherwise we have a product of divison algebras. W.l.o.g. $n_1 > 1$.

6.1 Simple algebras ($k = 1$)

Christandl and Zuiddam [12] prove a lower bound for the case of a single matrix algebra with entries from the ground field $\mathbb{F}$ using flattening. When d is odd, then their bound n^{d+1} is optimal. It is rather easy to generalize the flattening approach to semisimple algebras. We start with simple algebras that are not division algebras.

▶ **Theorem 12.** *Let $n > 1$ and let D be a division algebra of dimension ℓ, and let d be odd. Then we have an exponential lower bound of*

$$R(M_n(D), d) \geq \ell \cdot n^{d+1}.$$

Proof. Let $f_1 = 1, f_2, \ldots f_\ell$ be a basis for the division algebra D. Let $f_{i,j,\lambda}$ be the matrix in $M_n(D)$ that has the element f_λ in position (i, j) and zeros elsewhere. By the choice of f_1, $f_{i,j,1} = e_{i,j}$, where $e_{i,j}$ is the matrix that has a 1 in position (i, j) and zeros elsewhere. The set of all such $f_{i,j,\lambda}$ forms a basis of $M_n(D)$. The multiplication tensor in this case looks like

$$\sum_{i_1,\ldots,i_{d+1}\in[n],\lambda_1,\ldots,\lambda_d\in[\ell]} f^*_{i_1,i_2,\lambda_1} \otimes \ldots \otimes f^*_{i_d,i_{d+1},\lambda_d} \otimes (f_{\lambda_1} \cdot f_{\lambda_2} \cdot \ldots \cdot f_{\lambda_d}) f_{i_{d+1},i_1,1}.$$

The product $f_{\lambda_1} \cdot f_{\lambda_2} \cdot \ldots \cdot f_{\lambda_d}$ is an element of D, so it can be written as $\sum_{h=1}^{\ell} \alpha^h_{\lambda_1,\ldots,\lambda_d} f_h$ for suitable scalars $\alpha^h_{\lambda_1,\ldots,\lambda_d}$. Thus, the tensor can be rewritten as

$$\sum_{i_1,\ldots,i_{d+1}\in[n],\lambda_1,\ldots,\lambda_d\in[\ell]} f^*_{i_1,i_2,\lambda_1} \otimes \ldots \otimes f^*_{i_d,i_{d+1},\lambda_d} \otimes \sum_{h=1}^{\ell} \alpha^h_{\lambda_1,\ldots,\lambda_d} f_{i_{d+1},i_1,h}. \tag{2}$$

Now we flatten the multilinear map into a matrix. The tensor defines a multilinear map

$$D_1^{n\times n} \times D_2^{n\times n} \times \ldots \times D_d^{n\times n} \to D_{d+1}^{n\times n},$$

where we use the subscript i in D_i just to indicate the position. We will make it into a $((n^2\ell)^{(d+1)/2}) \times ((n^2\ell)^{(d+1)/2})$ matrix by combining the odd and even positions into the same parts, respectively. Therefore, the odd positions will index the rows and even will index the columns of the resulting matrix:

$$D_1^{n\times n} \otimes D_3^{n\times n} \otimes \ldots \otimes D_d^{n\times n} \to (D_2^{n\times n})^* \otimes (D_4^{n\times n})^* \otimes \ldots \otimes D_{d+1}^{n\times n}.$$

We further restrict this mapping to

$$D_1^{n\times n} \otimes \mathbb{F}^{n\times n} \otimes \ldots \otimes \mathbb{F}^{n\times n} \to (\mathbb{F}^{n\times n})^* \otimes (\mathbb{F}^{n\times n})^* \otimes \ldots \otimes D_{d+1}^{n\times n}.$$

The matrix is then of size $\ell n^{d+1} \times \ell n^{d+1}$. For every vector $f_{i_1,i_2,\lambda} \otimes f_{i_3,i_4,1} \otimes \cdots \otimes f_{i_d,i_{d+1},1} \in D_1^{n\times n} \otimes \mathbb{F}^{n\times n} \otimes \ldots \otimes \mathbb{F}^{n\times n}$, there is exactly one element in $(\mathbb{F}^{n\times n})^* \otimes (\mathbb{F}^{n\times n})^* \otimes \ldots \otimes D_{d+1}^{n\times n}$ such that the tensor product appears in Equation (2), namely, $f^*_{i_2,i_3,1} \otimes f^*_{i_4,i_5,1} \otimes \cdots \otimes f_{i_{d+1},i_1,\lambda}$. (Note that in this case, only one of the coefficients $\alpha^h_{\lambda_1,\ldots,\lambda_d}$ is nonzero, namely $\alpha^\lambda_{\lambda,1\ldots,1}$.) This means that the matrix is the identity matrix after appropriate permutation of rows and columns, which has rank $\ell \cdot n^{d+1}$.

The rank of the matrix is a lower bound on the rank of the tensor, which in turn is a lower bound on the multilinear complexity of $M_n(D)$. Thus, we have $R(M_n(D)) \geq \dim(D) \cdot n^{d+1}$. ◀

We also note that computing the rank of the flattened matrix for small ℓ, n and d using a computer algebra system gave us values suggesting the matrix rank to be $\ell \cdot n^{d+1}$. Therefore, a better lower bound using the flattening approach is unlikely.

6.2 Semisimple algebras

We can generalize the above proof for the case when the algebra $A = (D_1)^{n_1 \times n_1} \times (D_2)^{n_2 \times n_2} \times \ldots \times (D_k)^{n_k \times n_k}$.

▶ **Theorem 13.** *Let $A \cong \prod_{i=1}^{k} M_{n_i}(D_i)$ for arbitrary $n_i > 1$, arbitrary k, and D_i being a division algebra of dimension ℓ_i, and let d be odd. Then we have an exponential lower bound of*

$$R(A, d) \geq \sum_{i=1}^{k} \ell_i \cdot {n_i}^{d+1}.$$

Proof. The multiplication in A is a direct sum of multiplications in the k factors. The flattening matrix has a block structure when taking bases with respect to the k factors. Therefore, the flattening ranks add up. ◀

If d is even, we get a lower bound by using the fact that $R(A, d-1) \leq R(A, d)$. In the theorem above, all $n_i > 1$. The theorem also works in a similar way when $n_i = 1$ but D_i is noncommutative. In this case, we simply replace $\mathbb{F}$ by its algebraic closure (or just the splitting field of the division algebra). This will transform D_i into a matrix algebra with matrix size > 1.

6.3 Example

Consider the multiplication of d $n \times n$ matrices with elements from $D = \mathbb{F}[x]/(x^\ell + 1)$ with $x^\ell + 1$ being irreducible, i.e., we want to upper bound $R(M_n(D), d)$.

The degree of polynomials in the product matrix will be $d \cdot (\ell - 1)$, which means we will need $d \cdot (\ell - 1) + 1$ points to interpolate them. Thus, the algorithm is simply to evaluate the matrices at $d \cdot (\ell - 1) + 1$ points, calculate the result matrices of the product (complexity $R(M_n(\mathbb{F}), d)$) and interpolate at the points to get the final result matrix.

The number of multiplications will be $(d \cdot (\ell - 1) + 1) \cdot R(M_n(\mathbb{F}), d)$. We can do multiplication of d matrices in complexity n^{d+1} using the straight forward algorithm. Thus, it gives an upper bound of $R(M_n(D), d) \leq (d \cdot (\ell - 1) + 1) \cdot n^{d+1}$. We note that there is still a gap between our lower bound and this upper bound by a factor of d, but it is doubtful that the lower bound can be improved using flattening methods.

In particular, Theorem 8 can be viewed as the case of Theorem 13 when each $n_i = 1$ (the proof also works in this case). The flattening bound gives the lower bound ℓ while Theorem 8 gives $d\ell - (d - 1)$.

7 Upper triangular Matrices

In this section, we will look at the multilinear complexity of upper triangular matrices which are one of the special cases when A is noncommutative, but $A/\operatorname{rad}(A)$ is commutative. This is interesting as we saw in the case when $A/\operatorname{rad}(A)$ is noncommutative, we got exponential lower bounds in d, but for commutative algebras we saw linear lower bounds, and examples with linear upper bounds in d and polynomial upper bounds in general.

The result from Lemma 5 gives us a linear lower bound in d for general A. We will try to get more relevant lower and upper bounds for the special example of upper triangular matrices.

In the following, $U_n(\mathbb{F})$ denotes the algebra of upper triangular matrices. Its radical $\operatorname{rad}(U_n(\mathbb{F}))$ is the set of all upper triangular matrices with all diagonal elements 0. The quotient $U_n(\mathbb{F})/\operatorname{rad}(U_n(\mathbb{F}))$ will therefore be the set of all diagonal matrices, which is a commutative algebra and isomorphic to $\mathbb{F}^n$.

7.1 Lower Bound for $d \leq n$

We will use the flattening approach we used earlier to get a lower bound on the tensor rank for upper triangular matrices.

▶ **Theorem 14.** *For upper-triangular matrices, for $d \leq n$ odd, we have a lower bound of*

$$R(U_n(\mathbb{F}), d) \geq \binom{n+d}{n}.$$

Proof. We will be using flattening arguments similar to the one for general matrices. The multiplication tensor for the d-fold multiplication of upper triangular matrices looks as follows:

$$\mathbf{U}^{\otimes d} = \sum_{i_1, j_1, \ldots, i_{d+1}, j_{d+1} \in [n]} t_{i_1, j_1, \ldots, i_{d+1}, j_{d+1}} e^*_{i_1 j_1} \otimes \ldots \otimes e^*_{i_d j_d} \otimes e_{j_{d+1} i_{d+1}}$$

with $t_{i_1, j_1, \ldots, i_{d+1}, j_{d+1}} = (\prod_{k<d} \delta_{j_k i_{k+1}}) \cdot \delta_{j_d j_{d+1}} \cdot \delta_{i_{d+1} i_1}$ when $i_k \leq j_k$ for all $k \in [d]$, otherwise it will be 0. This means $t_{i_1, j_1, \ldots, i_{d+1}, j_{d+1}} = 1$ when $i_k = j_{k+1}$. From the upper triangular condition, we also have $i_k \leq j_k$.

We again flatten the tensor into a $(n^2)^{(d+1)/2} \times (n^2)^{(d+1)/2}$ matrix. As we saw in the proof of Theorem 12, each row or column has exactly one 1 and the matrix has full rank. With the extra condition of being upper triangular, we see that the rows will be non-zero if $i_1 \leq j_1, \ldots, i_d \leq j_d$, and as the columns are fixed with $j_1 = i_2 \leq j_2 = i_3 \ldots$, we basically get the rows with $i_1 \leq j_1 \leq i_3 \leq j_3 \leq \ldots \leq i_d \leq j_d$ will have exactly one 1. A similar thing can be done for the columns, with $j_{d+1} \leq i_2 \leq j_2 \leq i_4 \leq j_4 \leq \ldots \leq i_{d+1}$ with exactly one element 1. Therefore, the rank of the matrix will be the number of $i_1 \leq j_1 \leq i_3 \leq j_3 \leq \ldots \leq i_d \leq j_d$ sequences with $i, j \in [n]$. This number is $\binom{n+d}{n}$. ◀

7.2 Upper Bound for $d \gg n$

We see that unlike the case where $A/\operatorname{rad}(A)$ is noncommutative, the complexity flattens down as d becomes larger than n, becoming polynomial in d (for fixed n). We have the following upper bound:

▶ **Theorem 15.** *For multiplying upper-triangular matrices, for $d \gg n$, we have an upper bound of*

$$R(U_n(K), d) \leq O\left(\frac{(2d)^n}{\sqrt{n}}\right).$$

Proof. We can express the multiplication of d upper triangular matrices as

$$M_1 \cdot M_2 \cdot \ldots \cdot M_d.$$

Additionally, we can deconstruct any upper triangular matrix M_i as $M_i = D_i + N_i$, where D_i is a diagonal matrix and N_i has zeroes on the diagonal. We note that multiplying n upper triangular matrices which have zeroes on the diagonal yields 0, because they are in the radical. Then,

$$M_1 \cdot M_2 \cdot \ldots \cdot M_d = (D_1 + N_1)(D_2 + N_2), \ldots, (D_d + N_d).$$

We call the D_j's diagonal terms and the N_j's radical terms. We expand the product on the righthand side into a big sum. All the summands that have more than $n-1$ of the radical terms vanish. There are $\binom{d}{k}$ summands that have k radical terms, which in turn uniquely determines the positions of the diagonal terms in the summands. Therefore, there are $\sum_{k=0}^{n-1}\binom{d}{k}$ nonzero summands all together. We decompose each summand into rank-one tensors separately and in the trivial way. Each rank-one tensor in this decomposition corresponds to a subsequence of $(1,2,\dots n)$ of the form $(i_1, j_1, i_2, j_2, \dots, i_k, j_k)$ such that $i_1 < j_1 = i_2 < j_2 = j_3 < \dots = i_k < j_k$. The number of such subsequences is $\binom{n}{k+1}$ as choosing $i_1, i_2, \dots, i_k, j_k$ uniquely determines the subsequence. This sequence of indices corresponds to the basis elements e_{i_κ, j_κ} in each of the k radical terms. There is only one diagonal element that can stand between e_{i_κ, j_κ} and $e_{i_{\kappa+1}, j_{\kappa+1}}$, namely, e_{j_κ, j_κ}. (Recall that $j_\kappa = i_{\kappa+1}$.) Thus, each summand has a decomposition into rank-one tensors of length $\binom{n}{k+1}$.

Altogether, we have that

$$R(U_n(K), d) \le \sum_{k=0}^{n-1}\binom{d}{k}\binom{n}{k+1}.$$

We have $\binom{n}{n/2} \ge \binom{n}{k+1}$ for all $1 \le k \le n-1$ and we have that $\sum_{k=0}^{n-1}\binom{d}{k} \le d^n$. Thus, $R(U_n(K), d) \le \binom{n}{n/2} \cdot d^n$. Using Stirling's approximation, we have that

$$R(U_n(K), d) \le O\left(\frac{(2d)^n}{\sqrt{n}}\right). \qquad ◀$$

8 General algebras with commutative semisimple part

We see the above method can be used to give an upper bound on general algebras with commutative semisimple part, similar to the upper triangular matrices. We will only work over perfect fields (see [15] for a definition). Note that most fields are perfect, for instance, fields of characteristic zero, finite fields, and algebraically closed fields.

▶ **Theorem 16** (Wedderburn-Malcev Theorem, see [15]). *An algebra A over a perfect field can be written as $A = B \oplus \operatorname{rad}(A)$ with B being a subalgebra of A and $B \cong A/\operatorname{rad}(A)$.*

We see that in the situation of the theorem, we can write any element $x \in A$ as $x_B + x_{\operatorname{rad}(A)}$ with $x_B \in B$ and $x_{\operatorname{rad}(A)} \in \operatorname{rad}(A)$. This decomposition is unique. There are examples over non-perfect fields where such a B does not exist as a subalgebra of A, see [15].

The case when B is commutative is what is interesting to us, as the noncommutative case has an exponential lower bound anyway and we can remove the radical with Lemma 5. We focus on the case when $d \gg s$, where s is the smallest integer such that $(\operatorname{rad}(A))^s = \{0\}$. This number is also called the index of nilpotency. Obviously, $s \le \dim A$.

When we consider a product $x_1 \cdot x_2 \cdot \dots \cdot x_d$, we write it as $\prod_{i=1}^{d}(x_{i,B} + x_{i,\operatorname{rad}(A)})$. In this, terms with at least s terms from $\operatorname{rad}(A)$ will be 0. Therefore, when we expand the product as a sum, each nonzero summand contains at most $s-1$ factors from the radical.

Let ϕ^ℓ denote the $(\ell+1)$-linear map $B^\ell \times \operatorname{rad}(A) \to \operatorname{rad}(A)$, which takes $b_1, \dots, b_\ell \in B$ and $r \in \operatorname{rad}(A)$ and maps it to $b_1 \cdot \dots \cdot b_\ell \cdot r$. This is a restriction of the $(\ell+1)$-fold multiplication in A. We start with bounding the complexity of ϕ^ℓ.

▶ **Lemma 17.** *For all ℓ, $R(\phi^\ell) \le \ell^D D^{D+2}$, where D denotes the dimension of A.*

Proof. ϕ^1 is the bilinear multiplication $B \times \operatorname{rad}(A) \to \operatorname{rad}(A)$. We can simply bound this by $R(A, 2) \le D^2$, see Fact 4.

By Lemma 10, we have $R(B, \ell) \leq (\ell D)^D$. As seen above, $R(\phi^1) \leq D^2$. Consider a computation for the ℓ-fold multiplication in B, that is,

$$b_1 \cdot b_2 \cdot \ldots \cdot b_\ell = \sum_{i=1}^{r} F_{i,1}(b_1) F_{i,2}(b_2) \ldots F_{i,\ell}(b_\ell) \cdot w_i,$$

where $r \leq (\ell D)^D$. Furthermore, take a computation for ϕ^1,

$$b \cdot m = \sum_{j=1}^{r'} U_j(b) V_j(m) \cdot z_j$$

with $r' \leq D^2$. Plugging the first computation into the second, we get

$$\begin{aligned} b_1 \cdot \cdots \cdot b_\ell \cdot m &= \sum_{j=1}^{r'} U_j \left(\sum_{i=1}^{r} F_{i,1}(b_1) F_{i,2}(b_2) \ldots F_{i,\ell}(b_\ell) \cdot w_i \right) V_j(m) \cdot z_j \\ &= \sum_{j=1}^{r'} \sum_{i=1}^{r} F_{i,1}(b_1) F_{i,2}(b_2) \ldots F_{i,\ell}(b_\ell) U_j(w_i) V_j(m) \cdot z_j \\ &= \sum_{j=1}^{r'} \sum_{i=1}^{r} F_{i,1}(b_1) F_{i,2}(b_2) \ldots F_{i,\ell}(b_\ell) V_j(m) \cdot (U_j(w_i) z_j). \end{aligned}$$

Thus $R(\phi^\ell) \leq r r' = \ell^D D^{D+2}$. ◀

Concatenating k of the mappings ϕ_ℓ, we obtain an upper bound in a similar fashion to the upper triangular matrices.

▶ **Theorem 18.** *Let A be an algebra of dimension D over a perfect field $\mathbb{F}$ such that $A/\operatorname{rad} A$ is commutative. Then $R(A, d) \leq d^{(s-1)(D+1)} \cdot D^{(D+3)s}$, where s denotes the index of nilpotency of* $\operatorname{rad} A$*.*

Proof. Consider a product of d elements and write it as

$$x_1 \cdot x_2 \cdot \ldots \cdot x_d = \prod_{i=1}^{d} (x_{i,B} + x_{i,\mathrm{rad}(A)}).$$

We expand the product on the righthand side into a large sum. Summands with at least s factors from the radical will be zero by the definition of s. Let $k < s$. There are at most $\binom{d}{k}$ choices of k factors from the radical. As above, we will treat each summand separately. A summand is characterized by the k positions of the factors of the radical; they cut the product into $k+1$ parts of lengths $\ell_1, \ldots, \ell_{k+1}$ with $\ell_1 + \cdots + \ell_{k+1} = d - k$:

$$\begin{aligned} &x_{1,B} \cdots x_{\ell_1,B} \cdot x_{\ell_1+1,\mathrm{rad}(A)} x_{\ell_1+2,B} \cdots x_{\ell_1+\ell_2+2,B} \cdot x_{\ell_1+\ell_2+3,\mathrm{rad}(A)} \cdot x_{\ell_1+\ell_2+4,B} \cdots = \\ &\qquad \phi^{\ell_1}(x_{1,B}, \ldots, x_{\ell_1,B}, x_{\ell_1+1,\mathrm{rad}(A)}) \cdot \phi^{\ell_2}(x_{\ell_1+2,B}, \ldots, x_{\ell_1+\ell_2+2,B}, x_{\ell_1+\ell_2+3,\mathrm{rad}(A)}) \cdots \end{aligned}$$

(There are a total of $k+1$ factors on the righthand side, but we only wrote down the first two for the sake of readability.) For each ϕ^{ℓ_κ}, we can bound $R(\phi^{\ell_\kappa}) \leq \ell_\kappa^D D^{D+2}$ using Lemma 17. The last factor is a product of ℓ_{k+1} elements from B, we can simply bound the rank by $(\ell_{k+1} D)^D \leq \ell_{k+1}^D D^{D+2}$.

We trivially have $R(A, k+1) \leq D^{k+1}$, see Fact 4. By plugging computations for the ϕ^{ℓ_κ} in the computation for the $(k+1)$-fold multiplication in A, we get that there is a multilinear computation for our summand of length

$$\ell_1^D D^{D+2} \cdots \ell_{k+1}^D D^{D+2} \cdot D^{k+1} \leq d^{kD} \cdot D^{(D+3)(k+1)}.$$

Altogether we can bound the multilinear complexity by

$$\sum_{k=0}^{s-1} \binom{d}{k} d^{kD} \cdot D^{(D+3)(k+1)} \leq d^{s-1} d^{(s-1)D} \cdot D^{(D+3)s} = d^{(s-1)(D+1)} \cdot D^{(D+3)s}. \quad \blacktriangleleft$$

In the appendix, we present a more refined technique, that allows us to reduce the exponent of d in the upper bound.

9 Multilinear Alder–Strassen theorem

Finally, we prove the multilinear generalization of the Alder–Strassen theorem.

▶ **Theorem 19.** *Let A be a finite dimensional associative algebra with k maximal twosided ideals. Then $R(A, d) \geq d \cdot \dim A - (d-1)k$.*

Proof. We can assume $d \geq 3$, since the case $d = 1$ is trivial and the case $d = 2$ is the classical Alder–Strassen theorem.

By Lemma 5,

$$R(A, d) \geq R(A/\operatorname{rad} A, d) + d \cdot \dim \operatorname{rad} A. \tag{3}$$

Since A has k maximal twosided ideals, $A/\operatorname{rad} A$ has k maximal twosided ideals, too. and is of the form $A/\operatorname{rad} A = B_1 \times \cdots \times B_k$ with B_κ being simple algebras. Each $B_\kappa = M_{n_\kappa}(D_\kappa)$ with D_κ being a division algebra. Assume that $n_1 = \ldots n_j = 1$ and $n_{j+1}, \ldots, n_k > 1$. That means that $B_1, \ldots, B_j$ are division algebras.

We next prove that

$$R(A/\operatorname{rad} A, d) \geq d \cdot \dim(B_1 \times \cdots \times B_j) - (d-1)j + R(B_{j+1} \times \cdots \times B_k). \tag{4}$$

This is done by showing by induction that for all $1 \leq i \leq j$,

$$R(B_i \times \cdots \times B_k) \geq d \cdot \dim B_i - (d-1) + R(B_{i+1} \times \cdots \times B_k). \tag{5}$$

To prove Equation (5), we will inductively construct vector spaces $V_1, \ldots, V_d$ such that $V_\delta + B_i \times \{0\} \times \cdots \times \{0\} = B_i \times \cdots \times B_k$ for all $1 \leq \delta \leq d$, $(1, 0, \ldots, 0) \in V_\delta$ for $1 \leq \delta \leq d-1$, and

$$\begin{aligned} R(\phi) &\geq \delta \cdot \dim B_i - \delta + R(\phi|_{V_1 \times \cdots \times V_\delta \times (B_i \times \cdots \times B_k) \times \cdots \times (B_i \times \cdots \times B_k)}) \quad \text{for } 1 \leq \delta \leq d-1, \\ R(\phi) &\geq d \cdot \dim B_i - d + 1 + R(\phi|_{V_1 \times \cdots \times V_d}), \end{aligned} \tag{6}$$

where ϕ is the d-fold multiplication in $B_i \times \cdots \times B_k$. The proof is similar to the one of Theorem 8. Consider a multilinear computation for the d-fold multiplication in $B_i \times \cdots \times B_k$, i.e.,

$$x_1 \cdot x_2 \cdot \ldots \cdot x_d = \sum_{i=1}^{r} F_{i,1}(x_1) F_{i,2}(x_2) \ldots F_{i,d}(x_d) \cdot w_i.$$

Let $N = \dim B_i$. Similar to before, we can achieve that $F_{1,1}, \dots, F_{N,1}$ restricted to $B_i \times \{0\} \times \cdots \times \{0\}$ is a basis of $(B_i \times \{0\} \times \cdots \times \{0\})^*$. Let $a_1, \dots, a_N$ be the dual basis. Like before, we can assume that $a_N = (1, 0, \dots, 0)$ using sandwiching. We set $V_1 = \bigcap_{\nu=1}^{N-1} \ker F_{\nu,1}$. Then V_1 has the desired properties. Restricting to V_1 trivializes $N-1$ of the multilinear products, namely, the products $1, \dots, N-1$. Since $(1, 0, \dots, 0) \in V_1$, we can conclude that w.l.o.g. $F_{N-1,2}, \dots, F_{2N-1,2}$ restricted to $B_i \times \{0\} \times \cdots \times \{0\}$ form a basis and construct V_2 and so on. Since the induction stops at d, the dimension of the space V_d can be by one smaller, since we do not need $(1, 0, \dots, 0) \in V_d$. Therefore, restricting to V_d trivializes even N products and we obtain Equation (6).

$B_i \times \{0\} \times \cdots \times \{0\}$ is a twosided ideal in $B_i \times \cdots \times B_k$ with the property that $B_{i+1} \times \cdots \times B_k \cong B_i \times \cdots \times B_k / B_i \times \{0\} \times \cdots \times \{0\}$. Thus (5) follows from (6) in the similar way like in the end of the proof of Lemma 5.

Finally, we prove that

$$R(B_{j+1} \times \cdots \times B_k) \geq d \cdot \dim(B_{j+1} \times \cdots \times B_k). \tag{7}$$

This will finish the proof. To prove the last equation, we use Theorem 13 and get $R(B_{j+1} \times \cdots \times B_k) \geq \sum_{i=j+1}^{k} \dim D_i \cdot n_i^d$. (The exponent is d instead of $d+1$, since d might be even.) Since $\dim B_i = n_i^2 \dim D_i$, we are done when we can show that $n_i^d \geq d \cdot n_i^2$. The latter inequality is implied by $n_i^{d-2} \geq d$. Since $n_i \geq 2$, this is true when $d \geq 4$. In the case when $d = 3$, note that $n_i^{d-1} \geq d$ is sufficient, since d is odd. Combining (3), (4), and (7) finishes the proof. ◀

For $d \geq 3$, the proof in fact yields the stronger lower bound $d \cdot \dim A - (d-1)j$, where j is the number of division algebras in the decomposition of $A/\operatorname{rad} A$ into simple parts.

References

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

2 Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 522–539. SIAM, 2021. `doi:10.1137/1.9781611976465.32`.

3 Markus Bläser. Lower bounds for the bilinear complexity of associative algebras. *Comput. Complexity*, 9:73–112, 2000.

4 Markus Bläser. A complete characterization of the algebras of minimal bilinear complexity. *SIAM J. Comput.*, 34(2):277–298, 2004. `doi:10.1137/S0097539703438277`.

5 Markus Bläser. Fast matrix multiplication. *Theory Comput.*, 5:1–60, 2013. `doi:10.4086/toc.gs.2013.005`.

6 Markus Bläser and Bekhan Chokaev. Algebras of minimal multiplicative complexity. In *Proceedings of the 27th Conference on Computational Complexity, CCC 2012, Porto, Portugal, June 26-29, 2012*, pages 224–234. IEEE Computer Society, 2012. `doi:10.1109/CCC.2012.13`.

7 Markus Bläser and Christian Ikenmeyer. Introduction to geometric complexity theory. Lecture notes, 2018. URL: `https://pcwww.liv.ac.uk/~iken/teaching_sb/summer17/introtogct/gct.pdf`.

8 Nader H. Bshouty. Multilinear complexity is equivalent to optimal tester size. *Electron. Colloquium Comput. Complex.*, page 11, 2013. URL: `https://eccc.weizmann.ac.il/report/2013/011`, `arXiv:TR13-011`.

9 Nader H. Bshouty. Testers and their applications. In Moni Naor, editor, *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 327–352. ACM, 2014. `doi:10.1145/2554797.2554828`.

10 Harry Buhrman, Matthias Christandl, and Jeroen Zuiddam. Nondeterministic quantum communication complexity: the cyclic equality game and iterated matrix multiplication. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, volume 67 of *LIPIcs*, pages 24:1–24:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. `doi:10.4230/LIPIcs.ITCS.2017.24`.

11 Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic Complexity Theory*. Springer, 1997.

12 Matthias Christandl and Jeroen Zuiddam. Tensor surgery and tensor rank. *Comput. Complex.*, 28(1):27–56, 2019. `doi:10.1007/s00037-018-0164-8`.

13 P. M. Cohn. *Algebra*, volume 3. Wiley, 1991.

14 Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progression. *J. Symbolic Computation*, 9:251–280, 1990.

15 Yurij A. Drozd and Vladimir V. Kirichenko. *Finite Dimensional Algebras*. Springer, 1994.

16 François Le Gall. Powers of tensors and fast matrix multiplication. In Katsusuke Nabeshima, Kosaku Nagasaka, Franz Winkler, and Ágnes Szántó, editors, *International Symposium on Symbolic and Algebraic Computation, ISSAC '14, Kobe, Japan, July 23-25, 2014*, pages 296–303. ACM, 2014. `doi:10.1145/2608628.2608664`.

17 J. M. Landsberg. New lower bounds for the rank of matrix multiplication. *SIAM J. Comput.*, 43(1):144–149, 2014. `doi:10.1137/120880276`.

18 J. M. Landsberg. *Geometry and Complexity Theory*. Cambridge University Press, 2017.

19 Noam Nisan and Avi Wigderson. Lower bounds for arithmetic circuits via partial derivatives (preliminary version). In *36th Annual Symposium on Foundations of Computer Science, Milwaukee, Wisconsin, USA, 23-25 October 1995*, pages 16–25. IEEE Computer Society, 1995. `doi:10.1109/SFCS.1995.492458`.

20 Richard S. Pierce. *Associative Algebras*. Springer, 1982.

21 Ramprasad Saptharishi et al. A selection of lower bounds in arithmetic circuit complexity. Version 2021-07-27. URL: `https://github.com/dasarpmar/lowerbounds-survey/releases/tag/v9.0.3`.

22 Andrew J. Stothers. *On the complexity of matrix multiplication*. PhD thesis, The University of Edinburgh, 2010.

23 Volker Strassen. Gaussian elimination is not optimal. *Num. Math.*, 13:354–356, 1969.

24 Virginia Vassilevska Williams. Multiplying matrices faster than Coppersmith-Winograd. In *Proc. 44th Ann. ACM Symp. on Theory of Computing*, pages 887–898, 2012.

25 Shmuel Winograd. On multiplication of 2×2–matrices. *Lin. Alg. Appl.*, 4:381–388, 1971.

A Improvements of Theorem 18

Wedderburn's Theorem holds in a similar manner for modules over simple algebras. If A is an algebra, let $A^{n\times m}$ denote the vector space of all $n \times m$–matrices with entries from A.

▶ **Theorem 20** (Wedderburn). *Let A be a simple algebra with $A \cong M_n(D)$ for some division algebra D. For every A-left module $M \neq \{0\}$, there is a (unique) integer $m \geq 1$ such that M is isomorphic to $D^{n\times m}$.*

If C and D are algebras and M is a C-left module that is also a D-right module, then the module M is called a (C, D)*-bimodule*, if in addition $(am)b = a(mb)$ for all $a \in C$, $m \in M$, and $b \in D$. If $C = D$, M is also called a C-bimodule for short.

While the above bound of Theorem 18 is polynomial in d (when the dimension of the algebra is fixed), the algorithm can be improved and the exponent can be reduced. We sketch this approach below, but the actual bounds depend on a lot of parameters of the algebra, captured by the so-called path diagram of the algebra.

We decompose the semisimple algebra $B = B_1 \oplus \cdots \oplus B_n$ into simple algebras. (Here, the decomposition is written additively, and we consider the B_ν to be subspaces of B.) Each B_ν is a commutative division algebra. Let ℓ_ν be its dimension.

Let e_ν be the unit element of the algebra B_ν. It is well known [15], that $1 = e_1 + \cdots + e_n$ and that the $e_1, \ldots, e_n$ annihilate each other, that is, $e_\mu e_\nu = 0$ for all $\mu \neq \nu$. Write

$$A = 1 \cdot A \cdot 1 = (e_1 + \cdots + e_n)A(e_1 + \cdots + e_n) = \sum_{\mu,\nu=1}^{n} e_\mu A e_\nu.$$

Since $e_1, \ldots, e_n$ annihilate each other, the sum of vector spaces on the right hand side is direct. In the same way, we can decompose $\operatorname{rad} A$.

► **Fact 21.**
1. $\mathbb{F}[X] \otimes \mathbb{F}[Y] = \mathbb{F}[X, Y]$ *whenever X and Y are disjoint sets of variables.*
2. *Let $I \subseteq \mathbb{F}[X]$ and $J \subseteq \mathbb{F}[Y]$ be ideals. Then $F[X]/I \otimes \mathbb{F}[Y]/J = \mathbb{F}[X, Y](I + J)$.*

Since each B_i is commutative, we can write is as $\mathbb{F}[X_i]/I_i$ for some set of variables X_i and an ideal I_i. We assume that the sets of variables are pairwise disjoint.

The opposite algebra A^{opp} is the algebra obtained by "reversing" the multiplication of A. The multiplication $*$ in A^{opp} is defined by $a * b = b \cdot a$. When A is commutative, then A^{opp} and A are isomorphic (as algebras).

Let $R_{i,j} := e_i(\operatorname{rad} A)e_j$. The $R_{i,j}$ are B left-modules and B right-modules, so they are B-bimodules. However, the only nontrivial multiplication from the left is with elements from B_i and the only nontrivial multiplication from the right is with B_j. So it is more natural to view $R_{i,j}$ as a (B_i, B_j)-bimodule. A (B_i, B_j)-bimodule is isomorphic to a $B_i \otimes B_j^{\text{opp}}$-left module. Since B_j is commutative, this is in turn isomorphic to a $B_i \otimes B_j$ module. Take an element $m \in R_{i,j}$. $(B_i \otimes B_j)m$ is a submodule of $R_{i,j}$ and since $B_i \otimes B_j$ is a quotient of a polynomial ring, so is the submodule.

Let $D = \dim A$. There are $\binom{d}{k}$ choices for the factors from the radical, $0 \leq k \leq s$. For each factor of the radical, we choose a subspace R_{i_κ, j_κ}, $1 \leq \kappa \leq k$, such that $i_{\kappa+1} = j_\kappa$. So essentially, we sum over all path in the so-called path diagram of the algebra, see [15], which describes the structure of the radical. Along such a path, by the above consideration, the multiplication can be simulated by a polynomial multiplication, so we can do the multiplication "in one stroke" and do not need to glue different pieces like we did in the proof above.