ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Complexity 19 (2003) 43–60

Journal of COMPLEXITY

http://www.elsevier.com/locate/jco

# On the complexity of the multiplication of matrices of small formats

Markus Bläser

*Institut für Theoretische Informatik, Universität zu Lübeck, Wallstr. 40, 23560 Lübeck, Germany*

Received 28 March 2002; accepted 10 June 2002

**Abstract**

We prove a lower bound of $2mn + 2n - m - 2$ for the bilinear complexity of the multiplication of $n \times m$-matrices with $m \times n$-matrices using the substitution method ($m \geqslant n \geqslant 3$). In particular, we obtain the improved lower bound of 19 for the bilinear complexity of $3 \times 3$-matrix multiplication.


## 1. Introduction

In the late 1960s, Strassen [17] discovered a bilinear algorithm for multiplying $2 \times 2$-matrices using only 7 essential multiplications instead of 8. Using this astonishing algorithm recursively, Strassen derived an algorithm for multiplying $n \times n$-matrices with $O(n^{\log_2 7}) = O(n^{2.808})$ arithmetic operations. A lot of effort has been spent on improving Strassen's upper bound, see for example [2,8,16,18]. The current "world record" is held by Coppersmith and Winograd [8]. They exhibit an algorithm with $O(n^{2.376})$ arithmetic operations. But the only algorithm which is of practical relevance (at least until today) is Strassen's algorithm [17]. In all other of the mentioned algorithms, the constants hidden in the $O$-notation are far too huge.

One way to obtain faster algorithms of practical relevance is to find a good bilinear algorithm for multiplying matrices of some small format. Since any bilinear algorithm for multiplying $2 \times 2$-matrices requires at least 7 essential multiplications [20], we have to look for another format. The most promising formats are probably $3 \times 3$-matrix multiplication and $4 \times 4$-matrix multiplication. The best bilinear

*E-mail address:* blaeser@tcs.mu-luebeck.de.

0885-064X/02/$ - see front matter 
PII: S0885-064X(02)00007-9

algorithm for multiplying $3 \times 3$-matrices known so far uses 23 essential multiplications [14]. This yields an algorithm for multiplying $n \times n$-matrices with $O(n^{\log_3 23}) = O(n^{2.858})$ arithmetic operations. To improve Strassen's algorithm, an algorithm with 21 or less essential bilinear multiplications is required. The currently best upper bound for $4 \times 4$-matrix multiplication follows by applying Strassen's algorithm two times. This yields the upper bound 49. Any improvement of this result immediately yields an algorithm with less than $O(n^{\log_2 7})$ arithmetic operations.

Investigating the bilinear complexity of the multiplication of matrices of some small format is an interesting and challenging problem, see e.g. [7, Problem 17.1] for the $3 \times 3$ case. The above considerations show that any improvement of the upper bound might yield a new and faster matrix multiplication algorithm of practical relevance. On the other hand, any strengthened lower bound sheds new light on the problem of matrix multiplication and helps to understand its nature.

Before discussing the above issues in more detail, let us first settle the model of computation. In the following, if $V$ is a vector space, let $V^*$ denote its dual space.

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

(1) A sequence $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ with $f_\rho \in U^*$, $g_\rho \in V^*$, and $w_\rho \in W$ is called a *bilinear computation* of length $r$ for $\phi$ if

$$\phi(u, v) = \sum_{\rho=1}^{r} f_\rho(u) g_\rho(v) w_\rho \quad \text{for all } u \in U,\ v \in V.$$

(2) The length of a shortest bilinear computation for $\phi$ is called the *bilinear complexity* or the *rank* of $\phi$ and is denoted by $R(\phi)$.

If we want to emphasize the underlying ground field $k$, we will sometimes write $R_k(\phi)$ instead of $R(\phi)$. If we allow that $f_\rho$ and $g_\rho$ are both elements from $(U \times V)^*$, we get quadratic computations. The length of a shortest quadratic computation for $\phi$ is called the *multiplicative complexity* of $\phi$ and is denoted by $C(\phi)$ or $C_k(\phi)$. Obviously, $C(\phi) \leqslant R(\phi)$ and it is not hard to see that $R(\phi) \leqslant 2C(\phi)$ for any bilinear mapping $\phi$. Since for the design of fast matrix multiplication algorithms, bilinear computations play the most important role, we will focus on the bilinear complexity in the following.

### 1.1. Previous bounds

Let in the following $\langle \ell, m, n \rangle : k^{\ell \times m} \times k^{m \times n} \to k^{\ell \times n}$ denote the multiplication of $\ell \times m$-matrices with $m \times n$-matrices. Asymptotically, the best lower bound for $n \times n$-matrix multiplication over arbitrary fields is

$$R(\langle n, n, n \rangle) \geqslant \tfrac{5}{2} n^2 - 3n,$$

see [3]. However, this bound does not give any good results for $n<8$. For smaller formats, we have

$$R(\langle \ell, m, n\rangle) \geqslant \ell m + mn + \ell - m + n - 3 \quad \text{for } n \geqslant \ell \geqslant 2, \tag{1}$$

see [4]. This bound even holds for the multiplicative complexity.

For $\langle 2,2,2\rangle$, (1) together with Strassen's algorithms yields $R(\langle 2,2,2\rangle) = 7$, see also [20] for the lower bound. De Groote [9] even shows that up to equivalence, there is only one bilinear computation of length 7 for $\langle 2,2,2\rangle$. So this case is well understood.

The next format to investigate is $\langle 2,2,3\rangle$. (Note that the rank of matrix multiplication is invariant under permutations, see e.g. Eq. (14.21) in [7], so it does not matter which of the three possibilities—$\langle 2,2,3\rangle$, $\langle 2,3,2\rangle$, or $\langle 3,2,2\rangle$—we consider.) Here, (1) yields $R(\langle 2,2,3\rangle) \geqslant 10$ opposed to the upper bound $R(\langle 2,2,3\rangle) \leqslant 11$ obtained by combining Strassen's algorithm with an ordinary matrix–vector multiplication. Over $GF(2)$, we even have $R_{GF(2)}(\langle 2,2,3\rangle) = 11$, see [12]. The upper bound of 11 gives an exponent of 2.895 which is worse than the exponent by Strassen's algorithm. Interestingly, we have $C(\langle 2,2,3\rangle) = 10$ over fields of characteristic distinct from two by virtue of (1) for the multiplicative complexity and Waksman's algorithm [19].

For the format $\langle 2,3,3\rangle$, (1) yields $R(\langle 2,3,3\rangle) \geqslant 14$, opposed to the upper bound $R(\langle 2,3,3\rangle) \leqslant 15$ by Hopcroft and Kerr [12]. The upper bound of 15 gives an exponent of 2.811, still inferior to Strassen's algorithm.

The next format is $\langle 3,3,3\rangle$. This format is of particular interest, since it is the first one for which the best lower and upper bounds known so far differ significantly. On the other hand, the situation is not hopeless. Eq. (1) yields $R(\langle 3,3,3\rangle) \geqslant 18$. On the other hand, Laderman [14] shows $R(\langle 3,3,3\rangle) \leqslant 23$. Johnson and McLoughlin [13] present further bilinear computations for $\langle 3,3,3\rangle$ of length 23 that are not equivalent to Laderman's computation.

This upper bound gives an exponent of 2.854. An upper bound of 21 would yield a favorable exponent of 2.772.

For $\langle 4,4,4\rangle$, we have $33 \leqslant R(\langle 4,4,4\rangle) \leqslant 49$, so there is currently not much hope of determining the exact value of $R(\langle 4,4,4\rangle)$.

### 1.2. New results

The main achievement of the present work is another step towards the determination of the value of $R(\langle 3,3,3\rangle)$ as asked for in [7, Problem 17.1]. More precisely, we prove the new lower bound

$$R(\langle 3,3,3\rangle) \geqslant 19$$

over arbitrary fields. We will prove this bound in Section 4. The above bound is a special case of the following bound:

$$R(\langle n,m,n\rangle) \geqslant 2mn + 2n - m - 2 \qquad \text{for } m \geqslant n \geqslant 3$$

which will be proven in Section 5. Compared with (1), this is an improvement by one. This might seem like a small improvement at a first glance, but for instance, the problem whether $R(\langle 3,3,3\rangle)$ equals 17 or is strictly greater than 17 had been open for over 20 years after the proof that $R(\langle 3,3,3\rangle)\geqslant 17$ by Brockett and Dobkin [6]. Unfortunately, we are only able to prove this new bound for the bilinear complexity and only for formats of the type $\langle n,m,n\rangle$ (instead of $\langle \ell,m,n\rangle$).

## 2. Lower bound techniques

In this section we compile some of the results which were used by Alder and Strassen [1] to prove their so-called Alder–Strassen bound. Their method is a refinement of the substitution method, which is due to Pan [15]. Beside the original paper of Alder and Strassen, Chapter IV.2 of [11] and Chapter 17 of [7] are excellent treatments of the method of Alder and Strassen. The term "separate" and the Extension Lemma are taken from there, but everything is also implicitly in the work of Alder and Strassen. Alder and Strassen consider quadratic computations and multiplicative complexity. Since bilinear computations and bilinear complexity are only special cases, their results transfer to bilinear computations and bilinear complexity at once. Because we are concerned with bilinear complexity in this work, we focus on bilinear complexity in this section and state all of the results for the bilinear complexity explicitly.

**Definition 2.** Let $U$, $V$, and $W$ be vector spaces over some field $k$ and $\beta=(f_1,g_1,w_1,\ldots,f_r,g_r,w_r)$ be a bilinear 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 $\beta$ *separates* $(U_1,V_1,W_1)$, if there are disjoint sets of indices $I,J\subseteq\{\rho\mid w_\rho\notin W_1\}$ such that

$$U_1\cap\bigcap_{i\in I}\ker f_i=\{0\}\quad\text{and}\quad V_1\cap\bigcap_{j\in J}\ker g_j=\{0\}.$$

The latter condition is equivalent to the condition that $(f_i|_{U_1})_{i\in I}$ and $(g_j|_{V_1})_{j\in J}$ generate the dual spaces $U_1^*$ and $V_1^*$, respectively. This insight immediately yields the following lower bound:

**Lemma 3.** *Let $U$, $V$, and $W$ be vector spaces over some ground field $k$ and let $\beta=(f_1,g_1,w_1,\ldots,f_r,g_r,w_r)$ be a bilinear 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 $\beta$ separates $(U_1,V_1,W_1)$. Then*

$$r\geqslant\dim U_1+\dim V_1+\#\{\rho\mid w_\rho\in W_1\}.$$

To achieve good lower bounds by means of Lemma 3, one has to find an optimal bilinear computation that separates a "large" triple. An important tool to solve this task is the following "Extension Lemma". If $T$ is a subset of some vector space over a field $k$, let in the following $\langle T\rangle$ denote its $k$-linear span.

**Lemma 4** (Alder and Strassen). *Let $U$, $V$, and $W$ be vector spaces over a field $k$ and $\beta$ be a bilinear 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 $\beta$ separates the triple $(U_1, V_1, W_1)$. Then $\beta$ separates also $(U_2, V_1, W_1)$, or there is some $u \in U_2 \setminus U_1$ such that*

$$\phi(u, V) \subseteq \langle \phi(u, V_1) \rangle + W_1.$$

The Extension Lemma holds in the same manner for a subspace $V_2$ with $V_1 \subseteq V_2 \subseteq V$. If one replaces the term $\phi(u, V_1)$ by $\phi(U_2, V_1)$, then the Extension Lemma also holds for quadratic computations. For a proof in the quadratic case, we refer to [1,11] or [7]. For a proof in the bilinear case (with $\phi(u, V_1)$ instead of $\phi(U_2, V_1)$), see [5]. (Actually, all proofs in this paper also work with $\phi(U_2, V_1)$.)

## 3. Equivalence of computations

In this section, we establish some (well-known) equivalence transformations on the set of all computations of a given length $r$ for $\langle \ell, m, n \rangle$. We will exploit these equivalence relations in the following sections frequently. For a comprehensive theory of equivalence of computations for bilinear mappings, we refer to [10].

Let $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ be a bilinear computation of length $r$ for $\langle \ell, m, n \rangle$. Surely, *permuting the products* defines a equivalence relation. If $a \in k^{\ell \times \ell}$, $b \in k^{m \times m}$, and $c \in k^{n \times n}$ are invertible matrices, then

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

for all $x, y \in A$. Therefore, $\tilde{\beta} = (\tilde{f}_1, \tilde{g}_1 \tilde{w}_1, \ldots, \tilde{f}_r, \tilde{g}_r, \tilde{w}_r)$ is a bilinear computation for $\langle \ell, m, n \rangle$, where $\tilde{w}_\rho = a^{-1} w_\rho c$ and the linear forms $\tilde{f}_\rho$ and $\tilde{g}_\rho$ are defined by $\tilde{f}_\rho(x) = f_\rho(axb^{-1})$ and $\tilde{g}_\rho(y) = g_\rho(byc^{-1})$ for all $x, y$. Due to the shape of the above equation, this transformation is also called *sandwiching*. Finally, let $\bar{f}_\rho$ and $\bar{g}_\rho$ be defined by $\bar{f}_\rho(x) = g_\rho(x^\top)$ and $\bar{g}_\rho(y) = f_\rho(y^\top)$ for all $x \in k^{n \times m}$ and $y \in k^{m \times \ell}$, respectively. The computation $(\bar{f}_1, \bar{g}_1 w_1^\top, \ldots, \bar{f}_r, \bar{g}_r, w_r^\top)$ is a bilinear computation for $\langle n, m, \ell \rangle$, because

$$\sum_{\rho=1}^{r} \bar{f}_\rho(x) \bar{g}_\rho(y) w_\rho^\top = \left( \sum_{\rho=1}^{r} f_\rho(y^\top) g_\rho(x^\top) w_\rho \right)^\top = (y^\top x^\top)^\top = xy.$$

We denote the resulting "transposed" computation by $\beta^\top$.

## 4. Multiplying 3×3-matrices

In this section, we start with the proof of the new bound for the bilinear complexity of the multiplication of matrices of small formats. To be kind to the

reader's patience, we first prove the bound $R(\langle 3,m,3\rangle)\geqslant 5m+4$ for $m\geqslant 3$. In particular, the rank of $3\times 3$-matrix multiplication is at least 19. The rather technical and elaborate proof of the general bound is postponed to the next section. We will prove some intermediate results in its full generality, whenever we can achieve this with little extra effort.

By switching over to the algebraic closure, we may assume w.l.o.g. for the remainder of this paper that the underlying field is algebraically closed.

In the following, let $R^{e,h}$, $L_\eta^{e,h}$ for $0\leqslant\eta\leqslant h$, and $Z_{\eta'}^{e,h}$ for $1\leqslant\eta'\leqslant h$ denote the following subspaces of $k^{e\times h}$:

$$R^{e,h}=\begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0\\ * & \cdots & * & * & * & \cdots & *\\ \vdots & & \vdots & \vdots & \vdots & & \vdots\\ * & \cdots & * & * & * & \cdots & * \end{pmatrix},$$

$$L_\eta^{e,h}=\begin{pmatrix} 0 & \cdots & 0 & 0 & * & \cdots & *\\ 0 & \cdots & 0 & 0 & * & \cdots & *\\ \vdots & & \vdots & \vdots & \vdots & & \vdots\\ \underbrace{0 \quad \cdots \quad 0 \quad 0}_{\eta} & & & & * & \cdots & * \end{pmatrix},$$

$$Z_{\eta'}^{e,h}=\begin{pmatrix} 0 & \cdots & 0 & 0 & * & \cdots & *\\ 0 & \cdots & 0 & * & * & \cdots & *\\ \vdots & & \vdots & \vdots & \vdots & & \vdots\\ \underbrace{0 \quad \cdots \quad 0 \quad *}_{\eta'} & & & & * & \cdots & * \end{pmatrix}.$$

Each of the above three matrices denotes the vector space that is obtained by substituting each "$*$" by an arbitrary element from $k$. The extreme cases $L_0^{e,h}$ and $L_h^{e,h}$ are the whole space $k^{e\times h}$ and the nullspace, respectively. We have the inclusions $L_\eta^{e,h}\subset Z_\eta^{e,h}\subset L_{\eta-1}^{e,h}$ for all $1\leqslant\eta\leqslant h$. Furthermore, $R^{e,h}\cdot k^{h\times j}=R^{e,j}$ and $k^{e\times h}\cdot L_\eta^{h,j}=L_\eta^{e,j}$.

In the following, $m$ and $n$ always specify the format of the matrix multiplication map $\langle n,m,n\rangle$ we are investigating. Of crucial matter is the following lemma. It also holds for the more general case of the multiplication of $\ell\times m$-matrices with $m\times n$-matrices.

**Lemma 5.** *With the above notations, let $1\leqslant t\leqslant n$ and let $W_1,\ldots,W_t$ be subspaces of $k^{\ell\times n}$ such that $W_\tau\subseteq Z_\tau^{\ell,n}$ and $W_\tau\cap L_\tau^{\ell,n}=\{0\}$ for all $1\leqslant\tau\leqslant t-1$ as well as $W_t\subseteq L_t^{\ell,n}$ and $\dim W_t\leqslant\ell-1$. Then the following holds: if $\beta$ is a bilinear computation for $\langle\ell,m,n\rangle$, then $\beta$ separates the triple $(k^{\ell\times m},L_1^{m,n},W)$, where $W=W_1+\cdots+W_t$.*

**Proof.** As $W_\tau \cap L_\tau^{\ell,n} = \{0\}$ and $W_\tau \subseteq Z_\tau^{\ell,n} \subseteq L_{\tau-1}^{\ell,n}$, we may choose a projection $\pi_\tau : L_{\tau-1}^{\ell,n} \to L_\tau^{\ell,n}$ for all $1 \leqslant \tau < t$ such that $W_\tau \subseteq \ker \pi_\tau$. Let $p_\tau = \pi_\tau \circ \cdots \circ \pi_1$ for $1 \leqslant \tau < t$. For technical reasons, let $p_0$ be the identity. Clearly, $p_\tau(W) = W_{\tau+1} + \cdots + W_t$.

The proof of the lemma is divided into several steps, in each of these steps (except the first) we utilize the Extension Lemma to extend a given triple.

1. By definition, $\beta$ separates the triple $(\{0\}, \{0\}, W)$.

2. The computation $\beta$ separates $(\{0\}, L_t^{m,n}, W)$: if $L_t^{m,n} = \{0\}$, that is, $t = n$, then this has already been proven in step 1. Otherwise, assume that $\beta$ does not separate $(\{0\}, L_t^{m,n}, W)$. By the Extension Lemma, there is some $b \in L_t^{m,n} \setminus \{0\}$ such that

$$k^{\ell \times m} \cdot b \subseteq W.$$

The vector space $k^{\ell \times m} \cdot b$ on the left-hand side of the above inclusion is contained in $L_t^{\ell,n} \subset L_{t-1}^{\ell,n}$. Applying $p_{t-1}$ to the above inclusion yields $k^{\ell \times m} \cdot b \subseteq W_t$, a contradiction, since the dimension of the vector space $k^{\ell \times m} \cdot b$ is at least $\ell$ but $\dim W_t \leqslant \ell - 1$.

3. The computation $\beta$ separates $(R^{\ell,m}, L_t^{m,n}, W)$: otherwise, there is some $a \in R^{\ell,m} \setminus \{0\}$ such that

$$a \cdot k^{m \times n} \subseteq a \cdot L_t^{m,n} + W.$$

If we apply $p_{t-1}$ to this inclusion, we obtain $p_{t-1}(a \cdot k^{m \times n}) \subseteq L_t^{\ell,n} + W_t \subseteq L_t^{\ell,n}$. This is a contradiction, since $p_{t-1}(a \cdot k^{m \times n})$ contains a matrix that has a nonzero entry in its $t$th column. (The last assertion is easily seen as follows: there is an index pair $(\lambda, \mu)$ such that $a$ has a nonzero entry in position $(\lambda, \mu)$. Let $u \in k^{m \times n}$ be the matrix that has a one in position $(\mu, t)$ and zeros elsewhere. The matrix $au$ has a nonzero entry in its $t$th column and is in $L_{t-1}^{\ell,n}$. Thus $p_{t-1}(au) = au$.)

4. If $\beta$ separates $(R^{\ell,m}, L_{\tau+1}^{m,n}, W)$, then also $(R^{\ell,m}, L_\tau^{m,n}, W)$: otherwise, there is some $b \in L_\tau^{m,n} \setminus L_{\tau+1}^{m,n}$ such that

$$k^{\ell \times m} \cdot b \subseteq R^{\ell,m} \cdot b + W.$$

The vector spaces $k^{\ell \times m} \cdot b$ and $R^{\ell,m} \cdot b$ are contained in $L_\tau^{\ell,n}$. Thus application of $p_\tau$ yields

$$k^{\ell \times m} \cdot b \subseteq R^{\ell,n} \cap L_\tau^{\ell,n} + W_{\tau+1} + \cdots + W_t.$$

This is a contradiction, since $k^{n \times n} \cdot b$ contains a matrix that has a nonzero entry in position $(1, \tau + 1)$ (this is seen as in step 3) while the vector space on the right-hand side is contained in $Z_{\tau+1}^{\ell,n}$.

5. By induction (steps 3 and 4), $\beta$ separates $(R^{\ell,m}, L_1^{m,n}, W)$.

6. Finally, $\beta$ separates $(k^{\ell \times m}, L_1^{m,n}, W)$: otherwise, we can find a matrix $a \in k^{\ell \times m} \setminus R^{\ell,m}$ such that

$$a \cdot k^{m \times n} \subseteq a \cdot L_1^{m,n} + W \subseteq L_1^{\ell,n} + W.$$

This is a contradiction, because the set on the left-hand side contains a matrix that has a nonzero entry in position $(1, 1)$ (this is seen as in the previous steps), but the set on the right-hand side is contained in $Z_1^{\ell,n}$. □

In the following, we will assume that there is a bilinear computation $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ of length $r = 2mn + 2n - m - 3$ for $\langle n, m, n \rangle$ where $m \geqslant n \geqslant 3$. We will then prove that this assumption leads to a contradiction. (Unfortunately, we are not able to give a direct proof, since we will utilize a transformation which does not work for larger values of $r$.) The first goal in the course of the proof is the following: transform $\beta$ via equivalence transformations in such a way that

(1) $g_1, \ldots, g_M$ is a basis of $(k^{n\times m})^*$ where $M = mn$ and
(2) there is an $a \in \bigcap_{\mu=M+1}^{2M+n-m-1} \ker f_\mu$ with $0 < \operatorname{rk} a < n$.
(In other words, $a$ is neither zero nor of full rank.)

An important ingredient of the proof of this first goal is the following lemma which is proven in [4, Section 4].

**Lemma 6.** *Let $k$ be an algebraically closed field. If $a_1, \ldots, a_{2n-2} \in k^{n\times n}$, then there are invertible matrices $u, v \in k^{n\times n}$ such that*

$$u \cdot a_1 \cdot v, \ldots, u \cdot a_{2n-2} \cdot v \in Z_1^{n,n}.$$

W.l.o.g. we may assume that $f_1, \ldots, f_M$ is a basis of $(k^{n\times m})^*$. By Lemma 6, we can achieve $w_{2M-m}, \ldots, w_r \in Z_1^{n,n}$. By a suitable renumbering of the products $1, \ldots, M$, we may assume that in addition $R^{n,m} \cap \bigcap_{\mu=1}^{M-m} \ker f_\mu = \{0\}$. Hence, for all $b \in k^{n\times m}$ there is a $u \in R^{n,m}$ such that

$$f_\mu(b) = f_\mu(u), \quad 1 \leqslant \mu \leqslant M - m. \tag{2}$$

Let $c$ be an arbitrary element from $(\bigcap_{\mu=M-m+1}^{2M-m-1} \ker g_\mu) \setminus \{0\}$. Let $b \in k^{n\times m}$ be arbitrary and $u \in R^{n,m}$ be such that (2) holds. Then

$$(b - u)c = \sum_{\rho=2M-m}^{r} f_\rho(b - u) g_\rho(c) w_\rho.$$

Since $uc \in R^{n,n}$, this yields

$$bc \in R^{n,n} + \langle w_{2M-m}, \ldots, w_r \rangle \subseteq Z_1^{n,n}.$$

Because $b$ is arbitrary, $c$ cannot have full rank. (Since $m \geqslant n$, if $x \in k^{n\times m}$ and $y \in k^{m\times n}$ have full rank, then the homomorphisms induced by $u \mapsto xu$ and $v \mapsto vy$ are both surjective.)

Let $d = \dim \langle g_{M-m+1}, \ldots, g_{2M-m-1} \rangle$ and let $i_1, \ldots, i_d$ be indices from the set $\{M - m + 1, \ldots, 2M - m - 1\}$ such that $g_{i_1}, \ldots, g_{i_d}$ form a basis of this vector space. Obviously, $d \leqslant M - 1$. Choose indices $i_{d+1}, \ldots, i_M \in \{1, \ldots, M - m, 2M - m, \ldots, r\}$ such that $g_{i_1}, \ldots, g_{i_M}$ form a basis of $(k^{m\times n})^*$. (This is indeed possible, because $g_1, \ldots, g_r$ generate $(k^{m\times n})^*$.) Let $y_1, \ldots, y_M$ denote the dual basis of $g_{i_1}, \ldots, g_{i_M}$. In particular, $y_M \in \bigcap_{\delta=1}^{d} \ker g_{i_\delta}$. By construction, the linear span of $g_{i_1}, \ldots, g_{i_d}$ equals that of $g_{M-m+1}, \ldots, g_{2M-m-1}$. Thus, we even have $y_M \in \bigcap_{\mu=M-m+1}^{2M-m-1} \ker g_\mu$. By the above consideration (for $c = y_M$), this implies that $y_M$ does not have full rank.

We permute the products of $\beta$ such that $i_\mu = \mu$ for $1 \leqslant \mu \leqslant M$. We have achieved so far: $g_1, \ldots, g_M$ form a basis of $(k^{m\times n})^*$ with dual basis $y_1, \ldots, y_M$ and $y_M$ does not have full rank. By exploiting sandwiching, we may assume that the first column of $y_M$ is nonzero. Permute the other $y$'s in such a way, that the first columns of $y_{M+n-m}, \ldots, y_M$ are linearly independent.

If now $\dim\langle f_{M+n-m}, \ldots, f_r\rangle \leqslant M-1$, then $(\bigcap_{\rho=M+n-m}^{r} \ker f_\rho)$ contains a matrix $a \neq 0$ and we have

$$a \cdot y_\mu = \sum_{\rho=1}^{r} f_\rho(a) g_\rho(y_\mu) w_\rho = 0, \quad M+n-m \leqslant \mu \leqslant M.$$

The matrix $a$ cannot have full rank, since each of its $n$ rows is orthogonal to the first column of each of the $m-n+1$ many $y_\mu$ with $M+n-m \leqslant \mu \leqslant M$. Thus $0 < \operatorname{rk} a < n$ and we have reached our first goal.

Otherwise, choose $j_1, \ldots, j_M \in \{M+n-m, \ldots, r\}$ such that $f_{j_1}, \ldots, f_{j_M}$ form a basis. We may permute the products of $\beta$ such that for all $1 \leqslant \lambda \leqslant M$, $j_\lambda$ is mapped to $\lambda$ and for all $1 \leqslant \mu \leqslant M+n-m-1$, $\mu$ is mapped to $M+\mu$. Then $f_1, \ldots, f_M$ is a basis and we have $y_M \in \bigcap_{\mu=M+1}^{2M+n-m-1} \ker g_\mu$, because we had $y_M \in \bigcap_{\mu=1}^{M-1} \ker g_\mu$ before this permutation. Now we exchange the $f$'s with the $g$'s by switching over to $\beta^\top = (\bar{f}_1, \bar{g}_1, w_1^\top, \ldots, \bar{f}_r, \bar{g}_r, w_r^\top)$ which is again a bilinear computation for $\langle n, m, n\rangle$. After this exchange, $\bar{g}_1, \ldots, \bar{g}_M$ form a basis and for $a = y_M^\top$, we have $1 \leqslant \operatorname{rk} a \leqslant n-1$ and $a \in \bigcap_{\mu=M+1}^{2M+n-m-1} \ker \bar{f}_\mu$. This finishes the proof of the following lemma.

**Lemma 7.** *Let $m \geqslant n \geqslant 3$. If $R(\langle n, m, n\rangle) \leqslant 2mn + 2n - m - 3$, then there is a bilinear computation $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ of length $r = 2mn + 2n - m - 3$ for $\langle n, m, n\rangle$ such that with $M = mn$, $g_1, \ldots, g_M$ form a basis of $(k^{m\times n})^*$ and there is an $a \in \bigcap_{\mu=M+1}^{2M+n-m-1} \ker f_\mu$ with $1 \leqslant \operatorname{rk} a \leqslant n-1$.*

Let $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ with $r = 2mn + 2n - m - 3$ be such a computation for $\langle n, m, n\rangle$. Let $y_1, \ldots, y_M$ denote the dual basis of $g_1, \ldots, g_M$. By sandwiching, we may assume that $a$ has the form

$$a = \begin{pmatrix} 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ \underbrace{0 \quad \ldots \quad 0}_{n-\operatorname{rk} a} & & & \underbrace{0 \quad \ldots \quad 1}_{\operatorname{rk} a} & & & 0 & \ldots & 0 \end{pmatrix}.$$

For all $1 \leqslant \mu \leqslant M$,

$$ay_\mu = \sum_{\rho=1}^{r} f_\rho(a) g_\rho(y_\mu) w_\rho \in \langle w_\mu, w_{2M+n-m}, \ldots, w_r \rangle. \tag{3}$$

By a suitable renumbering of $g_1, \ldots, g_M$ and therefore also of $y_1, \ldots, y_M$, we may assume w.l.o.g. that $ay_1, \ldots, ay_s$ form a basis of the vector space $S = a \cdot k^{m \times n}$. Clearly, $s = \dim S = n \cdot \operatorname{rk} a$. Eq. (3) yields

$$S \subseteq \langle w_1, \ldots, w_s, w_{2M+n-m}, \ldots, w_r \rangle =: S'.$$

Let $d = \dim S' - \dim S$. Since $r = 2M + 2n - m - 3$, $d \leqslant n - 2$. Choose indices $i_1, \ldots, i_d \in \{1, \ldots, s, 2M + n - m, \ldots, r\} =: I$ such that

$$S + \langle w_{i_1}, \ldots, w_{i_d} \rangle = \langle w_1, \ldots, w_s, w_{2M+n-m}, \ldots, w_r \rangle = S'. \tag{4}$$

We now distinguish two cases: $\operatorname{rk} a > 1$ and $\operatorname{rk} a = 1$. Since the remainder of the proof is rather technical and elaborate in its full generality, we here restrict ourselves to case $n = 3$. The general case is proven in the next section.

If $n = 3$, then the case distinction is $\operatorname{rk} a = 2$ or $\operatorname{rk} a = 1$. We have $M = 3m$ and the length of $\beta$ is $r = 5m + 3$. The constant $d$ in (4) is either zero or one. Furthermore, $I = \{1, \ldots, s, r\}$.

We first treat the case $\operatorname{rk} a = 2$. In this case, $S = R^{3,3}$. If $d = 0$, then also $S' = R^{3,3}$. If $d = 1$, then we can transform $w_{i_1}$ by sandwiching with column operations into

$$w_{i_1} \in \begin{pmatrix} 0 & 0 & * \\ * & * & * \\ * & * & * \end{pmatrix} \quad \text{and therefore also} \quad S' \subseteq \begin{pmatrix} 0 & 0 & * \\ * & * & * \\ * & * & * \end{pmatrix}.$$

Since this transformation is achieved by mapping $w_\rho \mapsto w_\rho c$ and $g_\rho \mapsto \tilde{g}_\rho$ with $\tilde{g}_\rho(y) = g_\rho(yc^{-1})$ for some $c \in k^{3\times 3}$, this does neither affect $a$ nor $S$. In addition, Eq. (4) still holds. As $S \subseteq S'$, $\dim S' - \dim(S' \cap L_1^{3,3}) = 2$ and $\dim S' - \dim(S' \cap L_2^{3,3}) = 4$. Let $j_1, \ldots, j_4 \in I = \{1, \ldots, 6, r\}$ be pairwise distinct such that

$$\langle w_{j_1}, w_{j_2} \rangle \cap L_1^{3,3} = \{0\} \quad \text{and} \quad \langle w_{j_1}, \ldots, w_{j_4} \rangle \cap L_2^{3,3} = \{0\}.$$

Set $W_1 = \langle w_{j_1}, w_{j_2} \rangle$. Let $p$ be a projection onto $L_1^{3,3}$ such that $W_1 \subset \ker p$. Set $W_2 = \langle p(w_{j_3}), p(w_{j_4}) \rangle$. Since $w_{j_3}, w_{j_4} \in L_1^{3,3} + \langle w_{j_1}, w_{j_2} \rangle$, $w_{j_3}, w_{j_4} \in W_1 + W_2$. By construction, $W_\tau \cap L_\tau^{3,3} = \{0\}$ and $W_\tau \subseteq Z_\tau^{3,3}$ for $\tau = 1, 2$ as well as $w_{j_1}, \ldots, w_{j_4} \in W_1 + W_2$. Since $\beta$ separates $(k^{3\times m}, L_1^{m,3}, W_1 + W_2)$ by Lemma 5 (with $t = 3$, set formally $W_3 = \{0\}$), we obtain $r \geqslant 3m + 2m + 4 = 5m + 4$ by Lemma 3. This contradicts the assumption $r = 5m + 3$. Thus, we have indeed $r \geqslant 5m + 4$ in this case.

We now treat the case $\operatorname{rk} a = 1$. As we did once before, we replace $w_\rho$ with $w_\rho^\top$ by "transposing" the computation $\beta$. After this, $S = L_2^{3,3}$. If $d = 1$, then $w_{i_1} \notin S = L_2^{3,3}$. After possibly exchanging the first with the second column, we may assume w.l.o.g. that $w_{i_1}$ has a nonzero entry in its first column, that is, $w_{i_1} \notin L_1^{3,3}$. This does not affect $S$, since the third column is not affected. Choose distinct indices $j_1, j_2 \in \{1, \ldots, r\}$ such that $\langle w_{j_1}, w_{j_2} \rangle \cap L_1^{3,3} = \{0\}$. (Note that $\langle w_1, \ldots, w_r \rangle = k^{3\times 3}$.) If $d = 1$, we choose

$j_1 = i_1$. Now choose distinct indices $j_3, j_4 \in I \setminus \{j_1, j_2\}$. This is possible, because $I = \{1, 2, 3, r\}$, in particular $\#I = 4$. Set $W_1 = \langle w_{j_1}, w_{j_2} \rangle$. Let $p$ be a projection onto $L_1^{3,3}$ such that $W_1 \subset \ker p$. Set $W_2 = \langle p(w_{j_3}), p(w_{j_4}) \rangle$. By construction, $W_1 \cap L_1^{3,3} = \{0\}$ and $W_2 \subset S$, since $w_{j_3}, w_{j_4} \in S + \langle w_{i_1} \rangle$ and $w_{i_1} \in \ker p$ (in the case $d = 1$) or $w_{j_3}, w_{j_4} \in S$ ($d = 0$). Clearly, $\dim W_2 \leqslant 2$. Like in the case $\operatorname{rk} a = 2$, $w_{j_1}, \ldots, w_{j_4} \in W_1 + W_2$ holds. By Lemma 5, $\beta$ separates $(k^{3\times m}, L_1^{m,3}, W_1 + W_2)$ and by Lemma 3, $r \geqslant 3m + 2m + 4 = 5m + 4$. This again contradicts the assumption $r = 5m + 3$. Thus, also in this case we have indeed $r \geqslant 5m + 4$. This completes the proof of the following proposition.

**Proposition 8.** *For any field $k$ and for all $m \geqslant 3$,*

$$R(\langle 3, m, 3 \rangle) \geqslant 5m + 4.$$

In particular, we obtain the new lower bound 19 for the rank of $3 \times 3$-matrix multiplication.

**Corollary 9.** *For any field $k$, $R(\langle 3, 3, 3 \rangle) \geqslant 19$.*

## 5. The general case

For the proof of the new bound in the general case, we need the following technical lemmata. In the following, $G_h(e, k) \subset k^{e\times e}$ and $G^h(e, k) \subset k^{e\times e}$ denote the multiplicative groups of all invertible matrices of the form

$$\begin{pmatrix} i_h & 0 \\ 0 & x \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y & 0 \\ 0 & i_{e-h} \end{pmatrix},$$

respectively, where $i_h$ and $i_{e-h}$ denote the identity matrices of size $h \times h$ and $(e - h) \times (e - h)$, respectively, and $x$ and $y$ are invertible matrices of size $(e - h) \times (e - h)$ and $h \times h$, respectively. Multiplication with a matrix in $G_h(e, k)$ from the right corresponds to column operations of the columns $h + 1, \ldots, e$. In the same way, multiplication with a matrix in $G^h(e, k)$ from the left corresponds to row operations of the rows $1, \ldots, h$. (Of course, the same holds if we exchange right with left and columns with rows.) Let $P_h(e, k) \subset G_h(e, k)$ and $P^h(e, k) \subset G^h(e, k)$ denote the subgroups of all permutation matrices. These are the groups of all matrices where $x$ and $y$ are permutation matrices.

Most parts of the following proof work also for general matrix multiplication maps $\langle \ell, m, n \rangle$ (with possibly $\ell \neq n$). More precisely, we could state the following lemmata for $L_h^{\ell,n}$ instead of $L_h^{n,n}$ and so on—with the exception of Lemma 13. For Lemma 13, we would need the further assumption that $\ell \leqslant n$. Since we might interchange $\ell$ and $n$ in the course of the proof of Lemma 7, we cannot assure the condition $\ell \leqslant n$ in general. Therefore, we state everything only for the case $\ell = n$.

**Lemma 10.** *Let* $p : L_h^{n,n} \to L_{h+1}^{n,n}$ *be a projection and let* $W = \ker p$. *If* $u \in k^{n\times n}$ *is invertible and* $v \in G_{h+1}(n,k)$, *then there is a projection* $\pi : L_h^{n,n} \to L_{h+1}^{n,n}$ *such that* $\ker \pi = uWv$ *and* $\pi(uxv) = up(x)v$ *for all* $x \in L_h^{n,n}$.

**Proof.** We proceed in three steps.

1. Let $y \in L_h^{n,n}$ be arbitrary. The first $h+1$ columns of $yv$ equal the first $h+1$ columns of $y$, since multiplication with elements in $G_{h+1}(n,k)$ does not affect the first $h+1$ columns. Thus, $uyv \in L_h^{n,n}$. If in addition $y \in L_{h+1}^{n,n}$, then even $uyv \in L_{h+1}^{n,n}$.

2. If $w_1, \ldots, w_n$ is a basis of $W$ and $q_v$ denotes the $(h+1)$th column of $w_v$, then $uq_v$ is the $(h+1)$th column of $uw_vv$ by step 1. Since $q_1, \ldots, q_n$ are linearly independent, so are $uq_1, \ldots, uq_n$. Thus $(uWv) \cap L_{h+1}^{n,n} = \{0\}$. Since $(uWv) \subseteq L_h^{n,n}$ by step 1, we may choose $\pi$ as the projection along $uWv$ onto $L_{h+1}^{n,n}$.

3. If $x \in L_h^{n,n}$, then $x$ can be decomposed in a unique way into $x = w + y$ with $w \in W$ and $y \in L_{h+1}^{n,n}$ Thus, $up(x)v = uyv$. On the other hand, $uxv = uwv + uyv$. As $uwv \in uWv$ and $uyv \in L_{h+1}^{n,n}$, $\pi(uxv) = uyv = up(x)v$. □

**Lemma 11.** *Let* $V$ *be a vector space and* $X \subseteq Y$ *be subspaces of* $V$. *Let* $q$ *be a projection such that* $((\ker q) \cap Y) + X = Y$. *Then also* $((\ker q) \cap Y) + q(X) = Y$.

**Proof.** Replace $X$ by a subspace $X' \subseteq X$ such that $((\ker q) \cap Y) \oplus X' = Y$. Each $y \in Y$ can (uniquely) be written as $y = a + b$ with $a \in (\ker q) \cap Y$ and $b \in (\operatorname{im} q) \cap Y$. On the other hand, $b$ can be (uniquely) written as $b = c + x$ with $c \in (\ker q) \cap Y$ and $x \in X'$. Since $q$ is a projection, $b = q(b) = q(x)$. Thus $y = a + q(x)$ yielding $Y = ((\ker q) \cap Y) \oplus q(X')$. But $q(X) = q(X')$ and the lemma follows. □

**Lemma 12.** *Let* $h \leqslant n-1$, $s \leqslant n-1-h$ *and* $x_1, \ldots, x_s \in L_h^{n,n}$. *Then there are a matrix* $v \in G_h(n,k)$, *a number* $t$ *with* $1 \leqslant t \leqslant s+1$ *and subspaces* $W_1, \ldots, W_t$ *with* $W_\tau \subseteq Z_{h+\tau}^{n,n}$ *and* $W_\tau \cap L_{h+\tau}^{n,n} = \{0\}$ *for* $1 \leqslant \tau \leqslant t-1$ *as well as* $W_t \subseteq L_{h+t}^{n,n}$ *and* $\dim W_t \leqslant n-1$ *such that* $\langle x_1 v, \ldots, x_s v \rangle = W_1 + \cdots + W_t$.

**Proof.** The proof is by backward induction in $h$:

1. If $h = n-1$, then $s = 0$. Thus we choose $t = 1$, $W_1 = \{0\}$ und $v$ as the identity matrix.

2. Assume that $h < n-1$. If still $s = 0$, then we make the same choice as in step 1.

3. Assume therefore $s > 0$. Since $s \leqslant n-1-h$, there is an invertible matrix $u \in G_h(n,k)$ such that $x_1u, \ldots, x_su \in Z_{h+1}^{n,n}$. (To see this, note that one simply has to produce an extra zero in position $(1, h+1)$, because the $x$'s are already in $L_h^{n,n}$.) If now $x_1u, \ldots, x_su \in L_{h+1}^{n,n}$, then we complete the proof by setting $v = u$, $t = 1$, and $W_1 = \langle x_1u, \ldots, x_su \rangle$.

4. Otherwise, let $s' = \dim\langle x_1u, \ldots, x_su\rangle - \dim(\langle x_1u, \ldots, x_su\rangle \cap L_{h+1}^{n,n})$. After a suitable renumbering, we may assume that $\langle x_1u, \ldots, x_{s'}u\rangle \cap L_{h+1}^{n,n} = \{0\}$. Let $p$ be a

projection from $L_h^{n,n}$ onto $L_{h+1}^{n,n}$ with $\langle x_1u, \ldots, x_{s'}u\rangle \subseteq \ker p$. By the induction hypothesis, there is a matrix $v' \in G_{h+1}(n,k)$, a number $t$ with $2 \leqslant t \leqslant s - s' + 2$ and $W_2, \ldots, W_t$ such that $W_\tau \subseteq Z_{h+\tau}^{n,n}$ and $W_\tau \cap L_{h+\tau}^{n,n} = \{0\}$ for all $2 \leqslant \tau \leqslant t-1$, $W_t \subseteq L_{h+t}^{n,n}$ and $\langle p(x_{s'+1}u)v', \ldots, p(x_su)v'\rangle = W_2 + \cdots + W_t$. (Note the index shift.)

5. Let $W_1 = \langle x_1uv', \ldots, x_{s'}uv'\rangle$. By Lemma 10 (where we choose $u = 1$ and $v = v'$), there is a projection $\pi$ from $L_h^{n,n}$ onto $L_{h+1}^{n,n}$ such that $W_1 \subseteq \ker \pi$ and $\pi(yuv') = p(yu)v'$ for all $y \in L_h^{n,n}$. (If $y \in L_h^{n,n}$, so is $yu$.)

6. With $v = uv'$, we have $\langle x_1v, \ldots, x_sv\rangle = W_1 + \cdots + W_t$ by Lemma 11. To see this, choose $Y = \langle x_1v, \ldots, x_sv\rangle$, $X = \langle x_{s'+1}v, \ldots, x_sv\rangle$, and $q = \pi$. Then $Y \cap (\ker q) = \langle x_1v, \ldots, x_{s'}v\rangle$, thus the premises of Lemma 11 are fulfilled. Therefore, $Y = (\ker \pi) \cap Y + \pi(X)$ by Lemma 11. But $(\ker \pi) \cap Y = W_1$ and $\pi(X) = \langle \pi(x_{s'+1}v), \ldots, \pi(x_sv)\rangle = \langle p(x_{s'+1}u)v', \ldots, p(x_su)v'\rangle = W_2 + \cdots + W_t$. Since we have $v' \in G_{h+1}(n,k)$ and $x_1u, \ldots, x_su \in Z_{h+1}^{n,n}$, also $W_1 \subseteq Z_{h+1}^{n,n}$ holds. □

For technical reasons, the number $t$ of vector spaces $W_1, \ldots, W_t$ in the next lemma may be zero. In this case, the sum $W_1 + \cdots + W_t$ denotes the nullspace.

**Lemma 13.** *Let $h \leqslant n-1$, $s \leqslant n-1-h$ and $x_1, \ldots, x_s \in L_h^{n,n}$ such that $\langle x_1, \ldots, x_s\rangle \cap L_{n-1}^{n,n} = \{0\}$. Then there are matrices $u \in G^{s+1}(n,k)$, $v \in P_h(n,k)$ with $L_{n-1}^{n,n} \cdot v = L_{n-1}^{n,n}$, a number $t$ with $0 \leqslant t \leqslant s$ and subspaces $W_1, \ldots, W_t$ with $W_\tau \subseteq Z_{h+\tau}^{n,n}$ and $W_\tau \cap L_{h+\tau}^{n,n} = \{0\}$ such that $\langle ux_1v, \ldots, ux_sv\rangle = W_1 + \cdots + W_t$.*

**Proof.** The proof is again by backward induction in $h$:

1. If $h = n-1$, then $s = 0$ and we choose $t = 0$.

2. Assume $h < n-1$. If $\langle x_1, \ldots, x_s\rangle = \{0\}$, we again choose $t = 0$.

3. Therefore assume $\dim\langle x_1, \ldots, x_s\rangle > 0$. Then there exists an index $j$ with $h+1 \leqslant j \leqslant n-1$ such that at least one of the $x$'s has a nonzero entry in column $j$. Let $v_0$ be the permutation matrix that exchanges the $j$th column with the $(h+1)$th column and keeps the other columns fixed. We have $v_0 \in P_h(n,k)$. Let $s' = \dim\langle x_1v_0, \ldots, x_sv_0\rangle - \dim(\langle x_1v_0, \ldots, x_sv_0\rangle \cap L_{h+1}^{n,n})$. Clearly, $s' > 0$. W.l.o.g. assume that $\langle x_1v_0, \ldots, x_{s'}v_0\rangle \cap L_{h+1}^{n,n} = \{0\}$.

4. There is a matrix $u_0 \in G^{s+1}(n,k)$ such that each of the entries in the positions $(1, h+1), \ldots, (s+1-s', h+1)$ of each of the matrices $u_0x_1v_0, \ldots, u_0x_sv_0$ is zero. (This is due to the fact that the dimension of the vector space spanned by the $(h+1)$th columns of $x_1v_0, \ldots, x_sv_0$ is $s'$.) Let $x'_\sigma = u_0x_\sigma v_0$ for all $1 \leqslant \sigma \leqslant s$.

5. We have $\langle x'_1, \ldots, x'_{s'}\rangle \cap L_{h+1}^{n,n} = \{0\}$. Let $p$ be a projection onto $L_{h+1}^{n,n}$ such that $\langle x'_1, \ldots, x'_{s'}\rangle \subseteq \ker p$. As $L_{n-1}^{n,n}v_0 = L_{n-1}^{n,n}$, $\langle x_1, \ldots, x_s\rangle \cap L_{n-1}^{n,n} = \{0\}$ implies $\langle x'_1, \ldots, x'_s\rangle \cap L_{n-1}^{n,n} = \{0\}$. By Lemma 11, $\langle p(x'_{s'+1}), \ldots, p(x'_s)\rangle \cap L_{n-1}^{n,n} = \{0\}$. (In order to see this, choose $Y = \langle x'_1, \ldots, x'_s\rangle$, $X = \langle x'_{s'+1}, \ldots, x'_s\rangle$, and $q = p$. We obtain $Y = ((\ker p) \cap Y) + p(X)$. In particular, $Y \cap L_{n-1}^{n,n} = \{0\}$ yields $p(X) \cap L_{n-1}^{n,n} = \{0\}$.)

6. By the induction hypothesis, there are matrices $u' \in G^{s-s'+1}(n,k)$ and $v' \in P_{h+1}(n,k)$ with $L_{n-1}^{n,n}v' = L_{n-1}^{n,n}$, a number $t$ with $1 \leqslant t \leqslant s - s' + 1$, and subspaces

$W_2, \ldots, W_t$ with $W_\tau \subseteq Z^{n,n}_{h+\tau}$ and $W_\tau \cap L^{n,n}_{h+\tau} = \{0\}$ for all $2 \leqslant \tau \leqslant t$ such that $\langle u'p(x'_{s'+1})v', \ldots, u'p(x'_s)v' \rangle = W_2 + \cdots + W_t$.

7. Let $W_1 = \langle u'x'_1v', \ldots, u'x'_{s'}v' \rangle$. By Lemma 10, $W_1 \cap L^{n,n}_{h+1} = \{0\}$ and there is a projection $\pi$ onto $L^{n,n}_{h+1}$ with $W_1 \subseteq \ker \pi$ such that $\pi(u'yv') = u'p(y)v'$ for all $y \in L^{n,n}_h$.

8. With $y \in L^{n,n}_h$, also $u_0yv_0 \in L^{n,n}_h$. Thus, we may replace $y$ with $u_0yv_0$ above. By Lemma 11, $\langle ux_1v, \ldots, ux_sv \rangle = W_1 + \cdots + W_t$ with $u = u'u_0$ and $v = v_0v'$. (Choose $Y = \langle ux_1v, \ldots, ux_sv \rangle$, $X = \langle ux_{s'+1}v, \ldots, ux_sv \rangle$, and $q = \pi$. We have $Y = (\ker \pi) \cap Y + \pi(X)$, as well as $(\ker \pi) \cap Y = \langle ux_1v, \ldots, ux_{s'}v \rangle$, and furthermore $\pi(X) = \langle \pi(ux_{s'+1}v), \ldots, \pi(ux_sv) \rangle = \langle u'p(x'_{s'+1})v', \ldots, u'p(x'_s)v' \rangle$.) As $u' \in G^{s-s'+1}(n,k)$ and each entry in the positions $(1, h+1), \ldots, (s-s'+1, h+1)$ of each of the $x'_1, \ldots, x'_s$ is zero, we also have $W_1 \subseteq Z^{n,n}_{h+1}$. □

After these preparatory lemmata, we are now able to refute the assumption that there is a bilinear computation of length $r = 2mn + 2n - m - 3$ for $\langle n, m, n \rangle$ (for $m \geqslant n \geqslant 3$). Assume that $\beta = (f_1, g_1, w_1, \ldots, f_r, g_r, w_r)$ is a bilinear computation that fulfills the assertions of Lemma 7. Since the considerations in Section 4 have been for general $n$ until the case distinction right after Eq. (4), we may start with this case distinction.

Again, we first treat the case $\operatorname{rk} a > 1$. By sandwiching with column operations, we may achieve $w_{i_1}, \ldots, w_{i_d} \in Z^{n,n}_1$, because $d \leqslant n-2$. Since $S \subseteq R^{n,n} \subset Z^{n,n}_1$ and $S' = S + \langle w_{i_1}, \ldots, w_{i_d} \rangle$, this implies $S' \subseteq Z^{n,n}_1$. Let $p_v : k^{n \times n} \to k^{n \times n}/L^{n,n}_v$ be the canonical projection for $1 \leqslant v \leqslant n$. Let $e = \dim p_1(S')$ and $h = \dim p_2(S')$. We have $2 \leqslant \operatorname{rk} a \leqslant e \leqslant n-1$. (The lower bound is due to the fact that $\operatorname{rk} a > 1$, the upper bound follows from $S' \subseteq Z^{n,n}_1$.) Choose indices $j_1, \ldots, j_e \in I$ (where $I = \{1, \ldots, \dim S, 2M + n - m, \ldots r\}$ as defined in Section 4) such that

$$\langle w_{j_1}, \ldots, w_{j_e} \rangle \cap L^{n,n}_1 = \{0\}.$$

Like in the case $n = 3$, we want to construct a vector space $W_1$ that contains the matrices $w_{j_1}, \ldots, w_{j_e}$. But to obtain the improved bound, at least $n-1$ of the $w$'s should be in $W_1$, so if $e < n-1$, simply setting $W_1 = \langle w_{j_1}, \ldots, w_{j_e} \rangle$ is not enough. (In the case $n = 3$, $e = n-1$ was automatically true because $e \geqslant \operatorname{rk} a \geqslant 2$.) If $e < n-1$, we have to add some more $w$'s to $W_1$: by (4), we may assume that after a suitable permutation, $j_1, \ldots, j_\alpha \notin \{i_1, \ldots, i_d\}$ and $j_{\alpha+1}, \ldots, j_e \in \{i_1, \ldots, i_d\}$, where $\alpha := \operatorname{rk} a$.

By sandwiching with row operations of the rows $1, \ldots, n-\alpha$, we may achieve

$$w_{j_{\alpha+1}}, \ldots, w_{j_e} \in \left.\begin{pmatrix} 0 & * & * & \ldots & * \\ 0 & * & * & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & * & * & \ldots & * \\ * & * & * & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ * & * & * & \ldots & * \end{pmatrix}\right\} n - e. \tag{5}$$

This does not affect $S$, because all matrices in $S$ have solely zeros in the rows $1, \ldots, n-\alpha$. We even have that each $w \in S'$ is contained in the vector space on the right-hand side of (5), because otherwise $e$ would not be the dimension of $p_1(S')$.

Choose $h_1, \ldots, h_{n-e-1} \in \{1, \ldots, r\} \setminus I$ such that

$$\langle w_{j_1}, \ldots, w_{j_e}, w_{h_1}, \ldots, w_{h_{n-e-1}} \rangle \cap L_1^{n,n} = \{0\}.$$

This is possible, because $w_1, \ldots, w_r$ generate $k^{n \times n}$. By sandwiching with row operations of the rows $1, \ldots, n-e$, we may achieve

$$w_{h_1}, \ldots, w_{h_{n-e-1}} \in Z_1^{n,n}.$$

This does not affect $S$. In addition, all $w \in S'$ are still of the form as depicted in (5).

Since $d \leqslant n-2$, we may transform $w_{i_1}, \ldots, w_{i_d}$ via sandwiching with column operations of the columns $2, \ldots, n$ into

$$w_{i_1}, \ldots, w_{i_d} \in \left( \begin{array}{ccccc} 0 & 0 & * & \ldots & * \\ 0 & * & * & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & * & * & \ldots & * \\ * & * & * & \ldots & * \\ \vdots & \vdots & \vdots & & \vdots \\ * & * & * & \ldots & * \end{array} \right) \Bigg\} n-e. \tag{6}$$

This does not affect $S$ and we still have $w_{h_1}, \ldots, w_{h_{n-e-1}} \in Z_1^{n,n}$, because the first column remains unchanged. Since $S' = S + \langle w_{i_1}, \ldots, w_{i_d} \rangle$, all $w \in S'$ are transformed into the form depicted in (6). By Lemma 10, we still have

$$\langle w_{j_1}, \ldots, w_{j_e}, w_{h_1}, \ldots, w_{h_{n-e-1}} \rangle \cap L_1^{n,n} = \{0\},$$

because column operations of the columns $2, \ldots, n$ correspond to the right multiplication with an element from $G_1(n,k)$. Recall that $h = \dim p_2(S')$. It holds $h - e \geqslant \operatorname{rk} a \geqslant 2$. Choose indices $j_{e+1}, \ldots, j_h \in I$ such that

$$\langle w_{j_1}, \ldots, w_{j_h} \rangle \cap L_2^{n,n} = \{0\}.$$

Then also

$$\langle w_{j_1}, \ldots, w_{j_h}, w_{h_1}, \ldots, w_{h_{n-e-1}} \rangle \cap L_2^{n,n} = \{0\}$$

as the images of $w_{h_1}, \ldots, w_{h_{n-e-1}}$ under the canonical projection onto the positions $(2,1), \ldots, (n-e,1)$ are linearly independent but the images of $w_{j_1}, \ldots, w_{j_h}$ are all zero. Let

$$W_1 = \langle w_{j_1}, \ldots, w_{j_e}, w_{h_1}, \ldots, w_{h_{n-e-1}} \rangle$$

and

$$W_2 = \langle p(w_{j_{e+1}}), \ldots, p(w_{j_h}) \rangle,$$

where $p$ denotes a projection onto $L_1^{n,n}$ fulfilling $W_1 \subseteq \ker p$. By construction, $W_v \cap L_v^{n,n} = \{0\}$ and $W_v \subseteq Z_v^{n,n}$ for $v = 1, 2$. In addition, $w_\rho \in W_1 + W_2$ holds for $n - 1 + h - e \geqslant n + 1$ distinct $\rho$'s. (Note that $h - e \geqslant \operatorname{rk} a \geqslant 2$.) (The fact that

$w_{j_1}, \ldots, w_{j_e}, w_{h_1}, \ldots, w_{h_{n-e-1}} \in W_1 + W_2$ again follows from Lemma 11: let $Y = \langle w_{j_1}, \ldots, w_{j_h}, w_{h_1}, \ldots, w_{h_{n-e-1}} \rangle$, $X = \langle w_{j_{e+1}}, \ldots, w_{j_h} \rangle$, and $q = p$. Then $Y \cap (\ker q) = W_1$, thus the assumption of the lemma is fulfilled by $X$, $Y$, and $q$. Therefore $W_1 + p(X) = W_1 + W_2 = Y$.)

Finally, let $q_1, \ldots, q_{n-3} \in I \setminus \{j_1, \ldots, j_h\}$ be pairwise distinct. (Note that $\#I = \alpha n + d$ and $h \leqslant 2\alpha + d$.) Let $p'$ be a projection of $L_1^{n,n}$ onto $L_2^{n,n}$ with $W_2 \subseteq \ker p'$. By Lemma 12 (with $s = n - 3$) there are a matrix $v \in G_n(2, k)$, a $t$ with $3 \leqslant t \leqslant n$ and vector spaces $W_3, \ldots, W_t$ such that $W_\tau \cap L_\tau^{n,n} = \{0\}$ and $W_\tau \subseteq Z_\tau^{n,n}$ for all $3 \leqslant \tau \leqslant t - 1$ as well as $W_t \subseteq L_t^{n,n}$, $\dim W_t \leqslant n - 1$, and

$$p'(p(w_{q_1}))v, \ldots, p'(p(w_{q_{n-3}}))v \in W_3 + \cdots + W_t.$$

By Lemma 10, $(W_2 v) \cap L_2^{n,n} = \{0\}$ and there is a projection $\pi' : L_1^{n,n} \to L_2^{n,n}$ fulfilling $W_2 v \subseteq \ker \pi'$ and $\pi'(xv) = p'(x)v$ for all $x \in L_1^{n,n}$.

In the same way, $(W_1 v) \cap L_1^{n,n} = \{0\}$ and there is a projection $\pi$ onto $L_1^{n,n}$ fulfilling $W_1 v \subseteq \ker \pi$ and $\pi(xv) = p(x)v$ for all $x \in k^{n \times n}$.

Since $v \in G_2(n, k)$, we have $W_1 v \subseteq Z_1^{n,n}$ and $W_2 v \subseteq Z_2^{n,n}$. Therefore, the spaces $W_1 v, W_2 v, W_3, \ldots, W_t$ fulfill the assumptions of Lemma 5. In addition, we have $w_{q_\nu} v \in W = W_1 v + W_2 v + W_3 + \cdots + W_t$ for $1 \leqslant \nu \leqslant n - 3$ by Lemma 11. (Choose $Y = \langle w_{j_1} v, \ldots, w_{j_h} v, w_{q_1} v, \ldots, w_{q_{n-3}} v \rangle$, $X = \langle w_{q_1} v, \ldots, w_{q_{n-3}} v \rangle$, and $q = \pi' \circ \pi$.)

Thus, $w_\rho v \in W$ for at least $2n - 2$ distinct $\rho$'s. Via sandwiching, we can replace $w_\rho$ by $w_\rho v$. Lemma 5 now asserts that $\beta$ separates $(k^{n \times n}, L_1^{n,n}, W)$ and by Lemma 3, $r \geqslant 2mn + 2n - m - 2$ follows, contradicting the assumption $r = 2mn + 2n - m - 3$.

Now we treat the case $\operatorname{rk} a = 1$. As in the previous Section 4, by "transposing" the computation $\beta$ we may switch over from $w_\rho$ to $w_\rho^\top$. After this transformation, $S = L_{n-1}^{n,n}$. Let again $p_v : k^{n \times n} \to k^{n \times n} / L_v^{n,n}$ denote the canonical projection. If $d > 0$, then we may achieve $e := \dim p_1(\langle w_{i_1}, \ldots, w_{i_d} \rangle) > 0$ by a permutation of the columns $1, \ldots, n - 1$. This does not affect $S = L_{n-1}^{n,n}$. If $d = 0$, then we set $e = 0$ in the subsequent considerations. After a suitable permutation of indices, we may assume that $\langle w_{i_1}, \ldots, w_{i_e} \rangle \cap L_1^{n,n} = \{0\}$. Choose indices $j_1, \ldots, j_{n-1-d} \notin I$ such that

$$\langle w_{i_1}, \ldots, w_{i_e}, w_{j_1}, \ldots, w_{j_{n-1-d}} \rangle \cap L_1^{n,n} = \{0\}.$$

Let $h = e + n - 1 - d = \#\{i_1, \ldots, i_e, j_1, \ldots, j_{n-1-d}\}$. Via sandwiching with row operations, we can achieve that

$$w_{i_1}, \ldots, w_{i_e}, w_{j_1}, \ldots, w_{j_{n-1-d}} \in \left( \begin{array}{cccc} 0 & * & \ldots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \ldots & * \\ * & * & \ldots & * \\ \vdots & \vdots & & \vdots \\ * & * & \ldots & * \end{array} \right) \Bigg\} n - h.$$

Once again, this does not affect $S$. Let $W_1 = \langle w_{i_1}, \ldots, w_{i_e}, w_{j_1}, \ldots, w_{j_{n-1-d}} \rangle$.

Choose a projection $p$ onto $L_1^{n,n}$ fulfilling $W_1 \subseteq \ker p$. Let $s = d - e = \#\{i_{e+1}, \ldots, i_d\}$. We have $s = n - h - 1$. Lemma 13 assures the existence of matrices $u \in G^{n-h}(n,k)$ and $v \in P_n(1,k)$ with $L_{n-1}^{n,n} v = L_{n-1}^{n,n}$, a $t$ with $2 \leqslant t \leqslant s = n - h - 1 \leqslant n - 2$ (if $d = 0$ then $e = 0$, thus $s = 0$) as well as subspaces $W_2, \ldots, W_t$ fulfilling $W_\tau \subseteq Z_\tau^{n,n}$ and $W_\tau \cap L_\tau^{n,n} = \{0\}$ for all $2 \leqslant \tau \leqslant t$ such that $up(w_{i_{e+1}})v, \ldots, up(w_{i_d})v \in W_2 + \cdots + W_t$.

By Lemma 10, $(uW_1v) \cap L_1^{n,n} = \{0\}$, and there is a projection $\pi$ onto $L_1^{n,n}$ such that $\pi(uyv) = up(y)v$ for all $y \in L_1^{n,n}$. Since multiplication with $u$ from the left only affects the rows $1, \ldots, n - h$ and multiplication from the right with $v$ leaves the first column fixed, we still have $uW_1v \subseteq Z_1^{n,n}$. Moreover, $uSv = S$ and

$$(uW_1v + W_2 + \cdots + W_t) \cap L_{n-2}^{n,n} = \{0\}.$$

(For the last statement, exploit the facts that $(uW_1v) \cap L_1^{n,n} = \{0\}$, $W_\tau \subseteq Z_\tau^{n,n}$, and $W_\tau \cap L_\tau^{n,n} = \{0\}$ for all $2 \leqslant \tau \leqslant t \leqslant n - 2$.)

Let $q$ be a projection onto $S = L_{n-1}^{n,n}$ such that $uW_1v + W_2 + \cdots + W_t \subseteq \ker q$. Choose pairwise distinct indices $h_1, \ldots, h_{n-1} \in I \setminus \{i_1, \ldots, i_d\}$ (note that $\#I = n + d$) and set $W_{t+1} = \langle q(uw_{h_1}v), \ldots, q(uw_{h_{n-1}}v) \rangle \subset L_{n-1}^{n,n}$. The subspaces $uW_1v$, $W_2, \ldots, W_{t+1}$ fulfill the premises of Lemma 5 and their sum contains $uw_{i_1}v, \ldots, uw_{i_d}v$, $uw_{j_1}v, \ldots, uw_{j_{n-1-d}}v$ and $uw_{h_1}v, \ldots, uw_{h_{n-1}}v$ by Lemma 11.

Thus $uw_\rho v \in W := uW_1v + W_2 + \cdots + W_{t+1}$ for at least $2n - 2$ different $\rho$'s. Utilizing sandwiching, we may replace $w_\rho$ by $uw_\rho v$. By Lemma 5, $\beta$ separates $(k^{n\times n}, L_1^{n,n}, W)$ and by Lemma 3, $r \geqslant 2M + 2n - m - 2$ contradicting the assumption that $r = 2M + 2n - m - 3$. This finally proves our main theorem.

**Theorem 14.** *For any field $k$ and for all $m \geqslant n \geqslant 3$,*

$$R(\langle n, m, n \rangle) \geqslant 2mn + 2n - m - 2.$$

## References


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

[2] D. Bini, M. Capovani, G. Lotti, F. Romani, $O(n^{2.7799})$ complexity for matrix multiplication, Inform. Process. Lett. 8 (1979) 234–235.

[3] M. Bläser, A $\frac{5}{2}n^2$-lower bound for the rank of $n \times n$-matrix multiplication over arbitrary fields, in: Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 1999, pp. 45–50.

[4] M. Bläser, Lower bounds for the multiplicative complexity of matrix multiplication, Comput. Complexity 8 (1999) 203–226.

[5] M. Bläser, Lower bounds for the bilinear complexity of associative algebras, Comput. Complexity 9 (2000) 73–112.

[6] R.W. Brockett, D. Dobkin, On the optimal evaluation of a set of bilinear forms, Linear Algebra Appl. 19 (1978) 207–235.

[7] P. Bürgisser, M. Clausen, M.A. Shokrollahi, Algebraic Complexity Theory, Springer, Berlin, 1997.

[8] D. Coppersmith, S. Winograd, Matrix multiplication via arithmetic progression, J. Symbolic Comput. 9 (1990) 251–280.

[9] H.F. de Groote, On the varieties of optimal algorithms for the computation of bilinear mappings: optimal algorithms for $2 \times 2$-matrix multiplication, Theoret. Comput. Sci. 7 (1978) 127–148.
[10] H.F. de Groote, On the varieties of optimal algorithms for the computation of bilinear mappings: the isotropy group of a bilinear mapping, Theoret. Comput. Sci. 7 (1978) 1–24.
[11] H.F. de Groote, Lectures on the Complexity of Bilinear Problems, Lecture Notes in Computer Science, Vol. 245, Springer, Berlin, 1986.
[12] J.E. Hopcroft, L.R. Kerr, On minimizing the number of multiplications necessary for matrix multiplication, SIAM J. Appl. Math. 20 (1971) 20–36.
[13] R.W. Johnson, A.M. McLoughlin, Noncommutative bilinear algorithms for $3 \times 3$ matrix multiplication, SIAM J. Comput. 15 (2) (1986) 595–603.
[14] J. Laderman, A noncommutative algorithm for multiplying $3 \times 3$-matrices using 23 multiplications, Bull. Amer. Math. Soc. 82 (1976) 180–182.
[15] V.Ya. Pan, Methods for computing values of polynomials, Russian Math. Surveys 21 (1966) 105–136.
[16] A. Schönhage, Partial and total matrix multiplication, SIAM J. Comput. 10 (1981) 434–455.
[17] V. Strassen, Gaussian elimination is not optimal, Numer. Math. 13 (1969) 354–356.
[18] V. Strassen, Relative bilinear complexity and matrix multiplication, J. Reine Angew. Math. 375/376 (1987) 406–443.
[19] A. Waksman, On Winograd's algorithm for inner products, IEEE Trans. Comput. C-19 (1970) 360–361.
[20] S. Winograd, On multiplication of $2 \times 2$-matrices, Linear Algebra Appl. 4 (1971) 381–388.