THEORY OF COMPUTING, Volume 11 (11), 2015, pp. 285–298
www.theoryofcomputing.org

New Lower Bounds for the Border Rank of Matrix Multiplication

Joseph M. Landsberg * Giorgio Ottaviani†

Received October 1, 2013; Revised December 12, 2014; Published August 6, 2015

Abstract: The border rank of the matrix multiplication operator for $\mathbf{n}\times\mathbf{n}$ matrices is a standard measure of its complexity. Using techniques from algebraic geometry and representation theory, we show the border rank is at least $2\mathbf{n}^2-\mathbf{n}$. Our bounds are better than the previous lower bound (due to Lickteig in 1985) of $3\mathbf{n}^2/2+\mathbf{n}/2-1$ for all $\mathbf{n}\geq 3$. The bounds are obtained by finding new equations that bilinear maps of small border rank must satisfy, i. e., new equations for secant varieties of triple Segre products, that matrix multiplication fails to satisfy.

ACM Classification: F.2.1

AMS Classification: 68Q17, 68Q25, 15A99

Key words and phrases: matrix multiplication complexity, border rank

1 Introduction and statement of results

Finding lower bounds in complexity theory is considered difficult. For example, chapter 14 of [1] is titled "Circuit lower bounds: Complexity theory's Waterloo." The complexity of matrix multiplication is roughly equivalent to the complexity of many standard operations in linear algebra, such as taking the determinant or inverse of a matrix. A standard measure of the complexity of an operation is the minimal size of an arithmetic circuit needed to perform it. The exponent of matrix multiplication ω is defined to be $\underline{\lim}_{\mathbf{n}}\log_{\mathbf{n}}$ of the arithmetic cost to multiply $\mathbf{n}\times\mathbf{n}$ matrices, or equivalently, $\underline{\lim}_{\mathbf{n}}\log_{\mathbf{n}}$ of the minimal number of multiplications needed [3, Propositions 15.1, 15.10]. Determining the complexity of matrix

*Supported by NSF grant DMS-1006353.
†Member of GNSAGA-INDAM.

 DOI: 10.4086/toc.2015.v011a011

multiplication is a central question of practical importance. We give new lower bounds for its complexity in terms of border rank. These lower bounds are used to prove further lower bounds for tensor rank in [12, 17, 18].

Let A,B,C be vector spaces, with dual spaces A^*,B^*,C^*, and let $T : A^* \times B^* \to C$ be a bilinear map. The *rank* of T is the smallest r such that there exist $a_1,\dots,a_r \in A$, $b_1,\dots,b_r \in B$, $c_1,\dots,c_r \in C$ such that $T(\alpha,\beta) = \sum_{i=1}^r a_i(\alpha)b_i(\beta)c_i$. The *border rank* of T is the smallest r such that T can be written as a limit of a sequence of bilinear maps of rank r. Let $\underline{\mathbf{R}}(T)$ denote the border rank of T. See [11] or [3] for more on the rank and border rank of tensors, especially the latter for their relation to other measures of complexity.

Let $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle} : \mathrm{Mat}_{\mathbf{m}\times\mathbf{n}} \times \mathrm{Mat}_{\mathbf{n}\times\mathbf{l}} \to \mathrm{Mat}_{\mathbf{m}\times\mathbf{l}}$ denote the matrix multiplication operator. One has (see, e. g., [3, Propositions 15.1, 15.10]) that $\omega = \underline{\lim}_{\mathbf{n}}(\log_{\mathbf{n}}\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}))$. Naïvely $\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}) \le \mathbf{mnl}$ via the standard algorithm. In 1969, V. Strassen [23] showed that $\mathbf{R}(M_{\langle 2,2,2\rangle}) \le 7$ and, as a consequence, $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \le \mathcal{O}(\mathbf{n}^{2.81})$. Further upper bounds have been derived since then by numerous authors, with the current record $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \le \mathcal{O}(\mathbf{n}^{2.3728639})$ [15]. In 1983 Strassen showed [24] that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 3\mathbf{n}^2/2$, and shortly thereafter T. Lickteig [16] showed $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 3\mathbf{n}^2/2+\mathbf{n}/2-1$. Since then no further general lower bound had been found (although it is now known that $\underline{\mathbf{R}}(M_{\langle 2,2,2\rangle}) = 7$, see [10, 8]), and a completely different proof (using methods proposed by Mulmuley and Sohoni for Geometric complexity theory) that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle})) \ge 3\mathbf{n}^2/2-2$ was given in [4].

Our results are as follows:

Theorem 1.1. *Let* $\mathbf{n} \le \mathbf{m}$*. For all* $\mathbf{l} \ge 1$

$$\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}) \ge \frac{\mathbf{nl}(\mathbf{n}+\mathbf{m}-1)}{\mathbf{m}}. \tag{1.1}$$

Corollary 1.2.

$$\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{l}\rangle}) \ge 2\mathbf{nl}-\mathbf{l}, \tag{1.2}$$

$$\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 2\mathbf{n}^2-\mathbf{n}. \tag{1.3}$$

For 3×3 matrices, the state of the art is now $15 \le \underline{\mathbf{R}}(M_{\langle 3,3,3\rangle}) \le 20$, the upper bound is due to Smirnov [22].

Remark 1.3. The best lower bound for the *rank* of matrix multiplication, was, until recently, $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 5\mathbf{n}^2/2-3\mathbf{n}$, due to to Bläser [2]. After this paper was posted on arXiv, and using Corollary 1.2, it was shown that $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 3\mathbf{n}^2-4\mathbf{n}^{3/2}-\mathbf{n}$ by Landsberg in [12] and then, pushing the same methods further, A. Massarenti and E. Raviolo showed $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 3\mathbf{n}^2-2\sqrt{2}\mathbf{n}^{3/2}-3\mathbf{n}$ in [17, 18].

Our bounds come from explicit equations that bilinear maps of low border rank must satisfy. These equations are best expressed in the language of tensors. Our method is similar in nature to the method used by Strassen to get his lower bounds—we find explicit polynomials that tensors of low border rank must satisfy, and show that matrix multiplication fails to satisfy them. Strassen found his equations via linear algebra—taking the commutator of certain matrices. We found ours using representation theory and algebraic geometry. (Algebraic geometry is not needed for presenting the results. For its role in our method see [14].) More precisely, in §2 we define, for every p, a linear map

$$(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})_A^{\wedge p} \colon \mathbb{C}^{\mathbf{nl}\binom{\mathbf{mn}}{p}} \to \mathbb{C}^{\mathbf{ml}\binom{\mathbf{mn}}{p+1}} \tag{1.4}$$

and we prove that

$$\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}) \geq \binom{\mathbf{mn}-1}{p}^{-1} \operatorname{rank}\left[(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})_A^{\wedge p}\right].$$

In order to prove Theorem 1.1, we specialize this map for a judiciously chosen p to a subspace where it becomes injective. The above-mentioned equations are the minors of the linear map $(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})_A^{\wedge p}$.

The map (1.4) is of interest in its own right, we discuss it in detail in §4. This is done with the help of representation theory—we explicitly describe the kernel as a sum of irreducible representations labeled by Young diagrams.

Remark 1.4. It is conjectured in the computer science community that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle})$ grows like $\mathcal{O}(\mathbf{n}^{2+\varepsilon})$ for any $\varepsilon > 0$. A truly significant lower bound would be a function that grew like $\mathbf{n}^2 h(\mathbf{n})$ where h is an increasing function. No super-linear lower bound on the complexity of any explicit tensor (or any computational problem) is known, see [1, 25].

From a mathematician's perspective, all known equations for secant varieties of Segre varieties that have a geometric model arise by translating multi-linear algebra to linear algebra, and it appears that the limit of this technique is roughly the "input size" $3\mathbf{n}^2$.

Remark 1.5. The methods used here should be applicable to lower bound problems coming from the *Geometric Complexity Theory* (GCT) introduced by Mulmuley and Sohoni [19], in particular to separate the determinant (small weakly skew circuits) from polynomials with small formulas (small tree circuits). More precisely, the method used in this paper is a special case of the *Young flattenings* introduced in [14] and described in §2.2. In that paper they were used to obtain lower bounds on the symmetric border rank of polynomials, but the same methods can be used to compare any two polynomials. Both the method of partial derivatives [5] and the method of shifted partial derivatives [9, 7] may be viewed as special types of Young flattenings. One includes a space of polynomials S^dV into a space of linear maps (found via representation theory), and then $P \in S^dV$ cannot be in the $\mathsf{GL}(V)$-orbit closure of $Q \in S^dV$ if the rank of the linear map associated to P is greater than the rank of the linear map associated to Q. While it is unlikely these methods will prove the main conjecture of GCT, we hope they will be able to improve the state of the art, and at the same time give further insight to the differences and similarities between the determinant and permanent.

Overview

In §2 we describe the new equations to test for border rank in the language of tensors. Theorem 1.1 is proved in §3. We discuss the kernel of the map (1.4) in section §4. This analysis should be useful for future work. We conclude in §5 with a review of Lickteig's method for purposes of comparison. An appendix (§A) with basic facts from representation theory that we use is included for readers not familiar with the subject.

Acknowledgments

We thank K. Mulmuley and A. Wigderson for discussions regarding the perspective of computer scientists, P. Bürgisser and A. Wigderson for help improving the exposition, J. Hauenstein with help with computer calculations, and M. Bläser for help with the literature.

2 The new equations

Let A,B,C be complex vector spaces of dimensions $\mathbf{a},\mathbf{b},\mathbf{c}$, with $\mathbf{b}\leq\mathbf{c}$, and with dual vector spaces A^*,B^*,C^*. Then $A\otimes B\otimes C$ may be thought of as the space of bilinear maps $A^*\times B^*\to C$.

The most naïve equations for border rank are the so-called *flattenings*. Given $T\in A\otimes B\otimes C$, consider T as a linear map $B^*\to A\otimes C$ and write T_B for this map. Then $\underline{\mathbf{R}}(T)\geq \operatorname{rank}(T_B)$ and similarly for the analogous T_A,T_C. The inequality is because flattenings are sub-additive, and if T has rank one, the rank of T_B is one, so if T has border rank r, the rank of T_B is at most r. The reason one obtains a lower bound on border rank is because the rank of a linear map is determined by taking minors, an algebraic operation.

2.1 Strassen's equations

Strassen's equations [24] may be understood as follows (see §2.2 for the geometric origin of this perspective). As described in [20], tensor T_B with Id_A to obtain a linear map $B^*\otimes A\to A\otimes A\otimes C$ and skew-symmetrize the $A\otimes A$ factor to obtain a map

$$T_A^{\wedge 1}: B^*\otimes A\to \Lambda^2 A\otimes C\,.$$

If T is generic, then one can show that $T_A^{\wedge 1}$ will have maximal rank, and if $T=a\otimes b\otimes c$ is of rank one, $\operatorname{rank}((a\otimes b\otimes c)_A^{\wedge 1})=\mathbf{a}-1$. To see this, expand $a=a_1$ to a basis $a_1,\dots,a_{\mathbf{a}}$ of A with dual basis $\alpha^1,\dots,\alpha^{\mathbf{a}}$ of A^*. Then

$$T_A^{\wedge 1}=\sum_i [\alpha^i\otimes b]\otimes[a_1\wedge a_i\otimes c]\,,$$

so the image is isomorphic to $(A/a_1)\otimes c$.

It follows that $\underline{\mathbf{R}}(T)\geq \operatorname{rank}(T_A^{\wedge 1})/(\mathbf{a}-1)$ because $\operatorname{rank}(T_1+T_2)_A^{\wedge 1}\leq \operatorname{rank}(T_1)_A^{\wedge 1}+\operatorname{rank}(T_2)_A^{\wedge 1}$ for any tensors $T_1,T_2\in A\otimes B\otimes C$ and thus if $\underline{\mathbf{R}}(T)=r$, then $\operatorname{rank}(T)_A^{\wedge 1}\leq \operatorname{rank}(T_0)r=(\mathbf{a}-1)r$ where T_0 is a rank one tensor. Thus the best bound one could hope for with this technique is up to $r=\mathbf{b}\mathbf{a}/(\mathbf{a}-1)$. The minors of size $r(\mathbf{a}-1)+1$ of $T_A^{\wedge 1}$ give equations for the tensors of border rank at most r in $A\otimes B\otimes C$. This is most effective when $\mathbf{a}=3$.

When $\mathbf{a}>3$, for each 3-plane $A'\subset A$, consider the restriction $T|_{A'\otimes B\otimes C}$ and the corresponding equations, to obtain equations for the tensors of border rank at most r in $A\otimes B\otimes C$ as long as $r\leq 3\mathbf{b}/2$. This procedure is called *inheritance* (see [11, §7.4.2]).

We consider the following generalizations: tensor T_B with $\mathrm{Id}_{\Lambda^p A}$ to obtain a linear map $B^*\otimes \Lambda^p A\to \Lambda^p A\otimes A\otimes C$ and skew-symmetrize the $\Lambda^p A\otimes A$ factor to obtain a map

$$T_A^{\wedge p}: B^*\otimes \Lambda^p A\to \Lambda^{p+1}A\otimes C\,. \tag{2.1}$$

To avoid redundancies, assume $\mathbf{b} \leq \mathbf{c}$ and $p \leq \lceil \mathbf{a}/2 \rceil - 1$. Then, if $T = a \otimes b \otimes c$ is of rank one,

$$\operatorname{rank}((a \otimes b \otimes c)_A^{\wedge p}) = \binom{\mathbf{a}-1}{p}.$$

To see this, compute

$$T_A^{\wedge p} = \sum [\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_p} \otimes b] \otimes [a_1 \wedge a_{i_1} \wedge \cdots \wedge a_{i_p} \otimes c],$$

to conclude the image is isomorphic to $\Lambda^p(A/a_1) \otimes c$.

In summary:

Theorem 2.1. *Interpret $T \in A \otimes B \otimes C$ as a linear map $B^* \to A \otimes C$ and let $T_A^{\wedge p} : B^* \otimes \Lambda^p A \to \Lambda^{p+1} A \otimes C$ be the map obtained by skew-symmetrizing $T \otimes \mathrm{Id}_{\Lambda^p A}$ in the $A \otimes \Lambda^p A$ factor. Then*

$$\underline{\mathbf{R}}(T) \geq \frac{\operatorname{rank} T_A^{\wedge p}}{\binom{\mathbf{a}-1}{p}}.$$

Proof. Let $r = \underline{\mathbf{R}}(T)$ and let $T_\varepsilon = \sum_{i=1}^r T_{\varepsilon,i}$ be such that $\mathbf{R}(T_{\varepsilon,i}) = 1$ and $\lim_{\varepsilon \to 0} T_\varepsilon = T$. Then

$$\operatorname{rank} T_A^{\wedge p} \leq \operatorname{rank}(T_\varepsilon)_A^{\wedge p} \leq \sum_{i=1}^r \operatorname{rank}(T_{\varepsilon,i})_A^{\wedge p} = r\binom{\mathbf{a}-1}{p}. \qquad \square$$

Remark 2.2. Alternatively, one can compute the rank using the vector bundle techniques of [14].

When this article was posted on arXiv, we only knew that the minors of size $r\binom{\mathbf{a}-1}{p} + 1$ of the maps $T_A^{\wedge p}$ gave nontrivial equations for tensors of border rank at most r in $A \otimes B \otimes C$ for $r \leq 2\mathbf{a} - \sqrt{\mathbf{a}}$. Then, in [13], it was shown they actually give nontrivial equations at least up to $2\mathbf{b} - 3$ (the maximum possible would be $2\mathbf{b} - 2$).

We record the following proposition which follows from Stirling's formula and the discussion above.

Proposition 2.3. *The equations for the variety of tensors of border rank at most r in $A \otimes B \otimes C$ obtained by taking minors of $T_A^{\wedge p}$ are of degree $r\binom{\mathbf{a}-1}{p} + 1$. In particular, when r approaches the upper bound $2\mathbf{b}$ and $p = \lceil \frac{\mathbf{a}}{2} \rceil - 1$, the equations are asymptotically of degree*

$$\sqrt{\frac{2}{\pi}} \frac{2^{\mathbf{a}}\mathbf{b}}{\sqrt{\mathbf{a}-1}}.$$

Theorem 1.1 is obtained by applying the inheritance principle to the case of an $(\mathbf{n}+\mathbf{m}-1)$-plane $A' \subset A = \mathbb{C}^{\mathbf{nm}}$.

2.2 Origin of the equations corresponding to minors of (2.1)

This subsection is not used in the proof of the main theorem. We work in projective space as the objects we are interested in are invariant under rescaling.

Let

$$\mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) \subset \mathbb{P}(A \otimes B \otimes C)$$

denote the Segre variety of rank one tensors and let $\sigma_r(\mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ denote its r-th secant variety, the variety of tensors of border rank at most r.

In [14] we introduced a generalization of flattenings, called *Young flattenings*, which in the present context is as follows: Irreducible polynomial representations of the general linear group $\mathsf{GL}(A)$ correspond to partitions $\pi = (\pi_1, \dots, \pi_{\mathbf{a}})$; see §A.1. Let $S_\pi A$ denote the corresponding $\mathsf{GL}(A)$-module. Consider representations $S_\pi A, S_\mu B, S_\nu C$, and the identity maps $\mathrm{Id}_{S_\pi A} \in S_\pi A \otimes S_\pi A^*$, etc. Then we may consider

$$T \otimes \mathrm{Id}_{S_\pi A} \otimes \mathrm{Id}_{S_\mu A} \otimes \mathrm{Id}_{S_\nu A} \in A \otimes B \otimes C \otimes S_\pi A \otimes S_\pi A^* \otimes S_\mu B \otimes S_\mu B^* \otimes S_\nu C \otimes S_\nu C^* \,.$$

We may decompose $S_\pi A \otimes A$ according to the Pieri rule (see §A.3) and project to one irreducible component, say $S_{\tilde{\pi}} A$, where $\tilde{\pi}$ is obtained by adding a box to π, and similarly for C, while for B we may decompose $S_\mu B^* \otimes B$ and project to one irreducible component, say $S_{\hat{\mu}} B^*$, where $\hat{\mu}$ is obtained by deleting a box from μ. The upshot is a tensor

$$T' \in S_{\tilde{\pi}} A \otimes S_\mu B \otimes S_{\tilde{\nu}} C \otimes S_\pi A^* \otimes S_{\hat{\mu}} B^* \otimes S_\nu C^*$$

which we may then consider as a linear map, e. g.,

$$T' : S_\pi A \otimes S_\mu B^* \otimes S_\nu C \to S_{\tilde{\pi}} A \otimes S_{\hat{\mu}} B^* \otimes S_{\tilde{\nu}} C$$

and rank conditions on T' may give border rank conditions on T.

Returning to the minors of (2.1), the minors of size $t+1$ of $T_A^{\wedge p}$ give modules of equations which are contained in

$$\Lambda^{t+1}(\Lambda^p A \otimes B^*) \otimes \Lambda^{t+1}(\Lambda^{p+1} A^* \otimes C^*) = \bigoplus_{\substack{|\mu|=t+1,\\ |\nu|=t+1}} S_\mu(\Lambda^p A) \otimes S_{\mu'} B^* \otimes S_\nu(\Lambda^{p+1} A^*) \otimes S_{\nu'} C^*. \tag{2.2}$$

Determining which irreducible submodules of (2.2) actually contribute nontrivial equations appears to be difficult.

3 Proof of Theorem 1.1

Let M, N, L be vector spaces of dimensions $\mathbf{m}, \mathbf{n}, \mathbf{l}$. Write $A = M \otimes N^*$, $B = N \otimes L^*$, $C = L \otimes M^*$, so $\mathbf{a} = \mathbf{mn}$, $\mathbf{b} = \mathbf{nl}$, $\mathbf{c} = \mathbf{ml}$. The matrix multiplication operator $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle}$ is $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle} = \mathrm{Id}_M \otimes \mathrm{Id}_N \otimes \mathrm{Id}_L \in A \otimes B \otimes C$. (See [11, §2.5.2] for an explanation of this identification.) Let $U = N^*$. Then

$$(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle})_A^{\wedge p} : L \otimes U \otimes \wedge^p (M \otimes U) \to L \otimes M^* \otimes \wedge^{p+1}(M \otimes U) \,. \tag{3.1}$$

This is just the identity map on the L factor, so we may write $M^{\wedge p}_{\mathbf{m},\mathbf{n},\mathbf{l}} = \psi_p \otimes \mathrm{Id}_L$, where

$$\psi_p : \Lambda^p(M\otimes U)\otimes U \to M^*\otimes \Lambda^{p+1}(M\otimes U)\,. \tag{3.2}$$

The essential idea is to choose a subspace $A' \subset M\otimes U$ on which the "restriction" of ψ_p becomes injective for $p=\mathbf{n}-1$. Take a vector space W of dimension 2, and fix isomorphisms $U\simeq S^{\mathbf{n}-1}W^*$, $M\simeq S^{\mathbf{m}-1}W^*$. Let A' be the $\mathsf{SL}(W)$-direct summand $S^{\mathbf{m}+\mathbf{n}-2}W^* \subset S^{\mathbf{n}-1}W^*\otimes S^{\mathbf{n}-1}W^* = M\otimes U$.

Recall that $S^\alpha W$ may be interpreted as the space of homogeneous polynomials of degree α in two variables. If $f\in S^\alpha W$ and $g\in S^\beta W^*$ (with $\beta\le\alpha$) then we can perform the contraction $g\cdot f\in S^{\alpha-\beta}W$. In the case $f=l^\alpha$ is the power of a linear form l, then the contraction $g\cdot l^\alpha$ equals $l^{\alpha-\beta}$ multiplied by the value of g at the point l, so that $g\cdot l^\alpha=0$ if and only if l is a root of g.

Consider the natural skew-symmetrization map

$$A'\otimes\Lambda^{\mathbf{n}-1}(A')\to\Lambda^{\mathbf{n}}(A')\,. \tag{3.3}$$

Because $\mathsf{SL}(W)$ is reductive, there is a unique $\mathsf{SL}(W)$-complement A'' to A', so the projection $M\otimes U\to A'$ is well defined. Compose (3.3) with the projection

$$M\otimes U\otimes\Lambda^{\mathbf{n}-1}(A')\to A'\otimes\Lambda^{\mathbf{n}}(A') \tag{3.4}$$

to obtain

$$M\otimes U\otimes\Lambda^{\mathbf{n}-1}(A')\to\Lambda^{\mathbf{n}}(A')\,. \tag{3.5}$$

Now (3.5) gives a map

$$\psi'_p : U\otimes\Lambda^{\mathbf{n}-1}(A')\to M^*\otimes\Lambda^{\mathbf{n}}(A')\,. \tag{3.6}$$

We claim (3.6) is injective. (Note that when $\mathbf{n}=\mathbf{m}$ the source and target space of (3.6) are dual to each other.)

Consider the transposed map $S^{\mathbf{m}-1}W^*\otimes\Lambda^{\mathbf{n}}S^{\mathbf{m}+\mathbf{n}-2}W\to S^{\mathbf{n}-1}W\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}W$. It is defined as follows on decomposable elements (and then extended by linearity):

$$g\otimes(f_1\wedge\cdots\wedge f_{\mathbf{n}})\mapsto\sum_{i=1}^{\mathbf{n}}(-1)^{i-1}g(f_i)\otimes f_1\wedge\cdots\hat{f_i}\cdots\wedge f_{\mathbf{n}}\,.$$

We show this dual map is surjective. Let

$$l^{\mathbf{n}-1}\otimes(l_1^{\mathbf{m}+\mathbf{n}-2}\wedge\cdots\wedge l_{\mathbf{n}-1}^{\mathbf{m}+\mathbf{n}-2})\in S^{\mathbf{n}-1}W\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}W$$

with $l_i\in W$. Such elements span the target so it will be sufficient to show any such element is in the image. Assume first that l is distinct from the l_i. Since $\mathbf{n}\le\mathbf{m}$, there is a polynomial $g\in S^{\mathbf{m}-1}W^*$ which vanishes on $l_1,\dots,l_{\mathbf{n}-1}$ and is nonzero on l. Then, up to a nonzero scalar, $g\otimes(l_1^{\mathbf{m}+\mathbf{n}-2}\wedge\cdots\wedge l_{\mathbf{n}-1}^{\mathbf{m}+\mathbf{n}-2}\wedge l^{\mathbf{m}+\mathbf{n}-2})$ maps to our element.

Since the image is closed (being a linear space), the condition that l is distinct from the l_i may be removed by taking limits.

Finally, $\psi'_p\otimes\mathrm{Id}_L\colon B^*\otimes\Lambda^{\mathbf{n}-1}A'\to C\otimes\Lambda^{\mathbf{n}}A'$ is the map induced from the restricted matrix multiplication operator.

To complete the proof of Theorem 1.1, observe that an element of rank one in $A' \otimes B \otimes C$ induces a map $B^* \otimes \Lambda^{\mathbf{n}-1} A' \to C \otimes \Lambda^{\mathbf{n}} A'$ of rank $\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}$.

By Lemma 3.1 below, the border rank of $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle}$ must be at least the border rank of $T' \in A' \otimes B \otimes C$, and by Theorem 2.1

$$\underline{\mathbf{R}}(T') \geq \frac{\dim B^* \otimes \Lambda^{\mathbf{n}-1}(A')}{\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}} = \mathbf{nl} \frac{\binom{\mathbf{n}+\mathbf{m}-1}{\mathbf{n}-1}}{\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}} = \frac{\mathbf{nl}\,(\mathbf{n}+\mathbf{m}-1)}{\mathbf{m}}\,.$$

This concludes the proof of Theorem 1.1.

Lemma 3.1. *Let $T \in A \otimes B \otimes C$, let $A = A' \oplus A''$ and let $\pi \colon A \to A'$ be the linear projection, which induces $\tilde{\pi} \colon A \otimes B \otimes C \to A' \otimes B \otimes C$. Then $\mathbf{R}(T) \geq \mathbf{R}(\tilde{\pi}(T))$ and $\underline{\mathbf{R}}(T) \geq \underline{\mathbf{R}}(\tilde{\pi}(T))$.*

Proof. If $T = \sum_{i=1}^{r} a_i \otimes b_i \otimes c_i$ then $\tilde{\pi}(T) = \sum_{i=1}^{r} \pi(a_i) \otimes b_i \otimes c_i$. □

Remark 3.2. If we let $B' = U$, $C' = M$, then in the proof above we are just computing the rank of $(T')_A^{\wedge p}$ where $T' \in A \otimes B' \otimes C'$ is $\mathrm{Id}_U \otimes \mathrm{Id}_M$. The maximal border rank of a tensor T in $\mathbb{C}^{\mathbf{mn}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{n}}$ is $\mathbf{mn}$ which occurs anytime the map $T : \mathbb{C}^{\mathbf{mn}*} \to \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{n}}$ is injective, so T' is a generic tensor in $A \otimes B' \otimes C'$, and the calculation of $\mathrm{rank}(\psi'_p)$ is determining the maximal rank of $(T')_A^{\wedge p}$ for a generic element of $\mathbb{C}^{\mathbf{mn}} \otimes \mathbb{C}^{\mathbf{n}} \otimes \mathbb{C}^{\mathbf{m}}$. Also note that the projection $A \to A'$, viewed as linear map $S^{\mathbf{n}-1} W^* \otimes S^{\mathbf{n}-1} W^* \to S^{\mathbf{m}+\mathbf{n}-2} W^*$ is just polynomial multiplication.

4 The kernel through representation theory

The purpose of this section is to show that there are nontrivial equations for tensors of border rank less than $2\mathbf{n}^2$ that matrix multiplication *does* satisfy.

4.1 The kernel as a module

Assume $\mathbf{b} \leq \mathbf{c}$, so $\mathbf{n} \leq \mathbf{m}$. For a partition π, let $\ell(\pi)$ denote the number of parts of π, i. e., the largest k such that $\pi_k > 0$, where we write $\pi = (\pi_1, \ldots, \pi_k)$. Let π' denote the conjugate partition to π. See §A.1 for the definition of $S_\pi U$.

Example 4.1. Consider the case $\mathbf{m} = \mathbf{n} = 3$, take $p = 4$. Let

$$\alpha_1 = \ydiagram{2,1,1}\,, \qquad \alpha_2 = \ydiagram{2,2}\,, \qquad \alpha_3 = \ydiagram{3,1}\,.$$

Note that $\alpha_1 = \alpha_3'$, $\alpha_2 = \alpha_2'$. Then (see §A.2)

$$\Lambda^4(M \otimes U) = (S_{\alpha_3} M \otimes S_{\alpha_1} U) \oplus (S_{\alpha_2} M \otimes S_{\alpha_2} U) \oplus (S_{\alpha_1} M \otimes S_{\alpha_3} U)\,.$$

Observe that (via the Pieri rule, see §A.3)

$$\ydiagram{2,1,1} \otimes \ydiagram{1} = \ydiagram{3,1,1} \oplus \ydiagram{2,2,1},$$

$$\ydiagram{2,2} \otimes \ydiagram{1} = \ydiagram{3,2} \oplus \ydiagram{2,2,1},$$

$$\ydiagram{3,1} \otimes \ydiagram{1} = \ydiagram{4,1} \oplus \ydiagram{3,2} \oplus \ydiagram{3,1,1}.$$

Among the seven summands on the right-hand side, only

$$\ydiagram{4,1}$$

does not fit in the 3×3 square. The kernel of $M_{3,3,\mathbf{l}}^{\wedge 4}$ in this case contains $L^* \otimes S_{2,1,1}M \otimes S_{4,1}U$, corresponding to

$$\pi = \ydiagram{3,1} \qquad \pi + (1) = \ydiagram{4,1}$$

which has dimension $\mathbf{l} \cdot 24 \cdot 3 = 72\mathbf{l}$, and one can verify, either by computer or by explicitly writing down highest weight vectors of each module, that this is the entire kernel.

Lemma 4.2. $\ker(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})_A^{\wedge p} \supseteq \oplus_\pi S_{\pi'} M \otimes S_{\pi+(1)} U \otimes L$*, where the summation is over partitions*

$$\pi = (\mathbf{m}, \nu_1, \ldots, \nu_{\mathbf{n}-1}),$$

where $\nu = (\nu_1, \ldots, \nu_{\mathbf{n}-1})$ *is a partition of* $p - \mathbf{m}$*,* $\nu_1 \leq \mathbf{m}$ *and*

$$\pi + (1) = (\mathbf{m}+1, \nu_1, \ldots, \nu_{\mathbf{n}-1}).$$

Proof. Write $M_{\mathbf{m},\mathbf{n},\mathbf{l}}^{\wedge p} = \psi_p \otimes \mathrm{Id}_L$, where $\psi_p : \Lambda^p(M \otimes U) \otimes U \to M^* \otimes \Lambda^{p+1}(M \otimes U)$. Such modules are contained in the kernel by Schur's lemma, as there is no corresponding module in the target for it to map to. $\square$

Remark 4.3. We expect that equality holds in Lemma 4.2.

5 Review of Lickteig's bound

For comparison, we outline the proof of Lickteig's bound. (Expositions of Strassen's bound are given in several places, e. g., [11, Chap. 3] and [3, §19.3].) It follows in three steps. The first combines two standard facts from algebraic geometry: for varieties $X, Y \subset \mathbb{P}V$, let $J(X,Y) \subset \mathbb{P}V$ denote the join of X and Y. Then $\sigma_{r+s}(X) = J(\sigma_r(X), \sigma_s(X))$. If $X = \mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is a Segre variety, then $\sigma_s(\mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) \subseteq \mathrm{Sub}_s(A \otimes B \otimes C)$, where

$$\mathrm{Sub}_s(A\otimes B\otimes C) := \left\{ T \in A\otimes B\otimes C \;\middle|\; \begin{array}{l} \exists A' \subset A,\ \ B' \subset B, C' \subset C, \\ \dim A' = \dim B' = \dim C' = s, \\ T \in A'\otimes B'\otimes C' \end{array} \right\}.$$

See, e. g., [11, §7.1.1] for details. Next Lickteig observes that if $T \in \sigma_{r+s}(\mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$, then there exist A', B', C' each of dimension s such that, thinking of $T : A^* \otimes B^* \to C$,

$$\dim\big(T((A')^\perp \otimes B^* + A^* \otimes (B')^\perp)\big) \le r. \tag{5.1}$$

This follows because the condition is a closed condition and it holds for points on the open subset of points in the span of $r+s$ points on $\mathrm{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$.

Finally, for matrix multiplication, with $A = M \otimes N^*$, etc., he defines $M' \subset M$, $N^{*\prime} \subset N^*$ to be the smallest spaces such that $A' \subseteq M' \otimes N^{*\prime}$ and similarly for the other spaces. Then one applies (5.1) combined with the observation that $M|_{(A')^\perp \otimes B^*} \subseteq M' \otimes L^*$, etc., and keeps track of the various bounds to conclude.

A Appendix: Facts from representation theory

A.1 Representations of $\mathsf{GL}(V)$

The irreducible representations of $\mathsf{GL}(V)$ are indexed by sequences $\pi = (p_1, \dots, p_l)$ of non-increasing integers with $l \le \dim V$. Those that occur in $V^{\otimes d}$ are partitions of d, and we write $|\pi| = d$ and $S_\pi V$ for the module. $V^{\otimes d}$ is also an $\mathfrak{S}_d$-module, and the groups $\mathsf{GL}(V)$ and $\mathfrak{S}_d$ are the commutants of each other in $V^{\otimes d}$ which implies the famous Schur-Weyl duality that

$$V^{\otimes d} = \bigoplus_{|\pi|=d, \ell(\pi)\le \mathbf{v}} S_\pi V \otimes [\pi]$$

as a $(\mathsf{GL}(V) \times \mathfrak{S}_d)$-module, where $[\pi]$ is the irreducible $\mathfrak{S}_d$-module associated to π. Repeated numbers in partitions are sometimes expressed as exponents when there is no danger of confusion, e. g., $(3,3,1,1,1,1) = (3^2, 1^4)$. For example, $S_{(d)}V = S^d V$ and $S_{(1^d)}V = \Lambda^d V$. The modules $S_{s^{\mathbf{v}}}V = (\Lambda^{\mathbf{v}} V)^{\otimes s}$ are trivial as $\mathsf{SL}(V)$-modules. The module $S_{(22)}V$ is the home of the Riemann curvature tensor in Riemannian geometry. See any of [11, Chap. 6], [6, Chap 6] or [21, Chap. 9] for more details on the representations of $\mathsf{GL}(V)$ and what follows.

A.2 Useful decomposition formulas

To decompose $S^2(A \otimes B)$ as a $\mathsf{GL}(A) \times \mathsf{GL}(B)$-module, note that given $P \in S^2 A$ and $Q \in S^2 B$, the product of P and Q defined by $P \otimes Q(\alpha \otimes \beta, \alpha' \otimes \beta') := P(\alpha, \alpha')Q(\beta, \beta')$ will be in $S^2(A \otimes B)$. Similarly, if $P \in \Lambda^2 A$ and $Q \in \Lambda^2 B$, $P \otimes Q$ will also be symmetric as

$$P(\alpha', \alpha)Q(\beta', \beta) = [-P(\alpha, \alpha')][-Q(\beta, \beta')] = P(\alpha, \alpha')Q(\beta, \beta').$$

Since the dimensions of these spaces add to the dimension of $S^2(A\otimes B)$ we conclude

$$S^2(A\otimes B)=(S^2A\otimes S^2B)\oplus(\Lambda^2A\otimes\Lambda^2B)\,.$$

By an analogous argument, we have the decomposition

$$\Lambda^2(A\otimes B)=(S^2A\otimes\Lambda^2B)\oplus(\Lambda^2A\otimes S^2B)\,.$$

More generally (see, e. g., [11, §6.5.2]) we have

$$\Lambda^p(A\otimes B)=\oplus_{|\pi|=p}S_\pi A\otimes S_{\pi'}B\,, \tag{A.1}$$

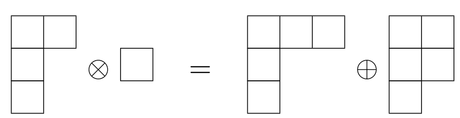
$$S^p(A\otimes B)=\oplus_{|\pi|=p}S_\pi A\otimes S_\pi B \tag{A.2}$$

where π' denotes the conjugate partition to π, that is, if we represent $\pi=(p_1,\dots,p_n)$ by a Young diagram, with p_j boxes in the j-th row, the diagram of π' is obtained by reflecting the diagram of π along the *NW* to *SE* axis.

A.3 The Pieri rule

The decomposition of $S_\pi V\otimes V$ is multiplicity free, consisting of a copy of each $S_\mu V$ such that the Young diagram of μ is obtained from the Young diagram of π by adding a box. (Boxes must be added in such a way that one still has a partition and the number of rows is at most the dimension of V.)

For example:



More generally, the *Pieri formula* states that $S_\pi V\otimes S^dV$ decomposes multiplicity free into the sum of all $S_\mu V$ that can be obtained by adding d boxes to the Young diagram of π in such a way that no two boxes are added to the same column, and $S_\pi V\otimes\Lambda^dV$ decomposes multiplicity free into the sum of all $S_\mu V$ that can be obtained by adding d boxes to the Young diagram of π in such a way that no two boxes are added to the same row. See any of the standard references given above for details.

References

[1] SANJEEV ARORA AND BOAZ BARAK: *Computational Complexity: A Modern Approach*. Cambridge Univ. Press, 2009. 285, 287

[2] MARKUS BLÄSER: A $\frac{5}{2}n^2$-lower bound for the rank of $n\times n$-matrix multiplication over arbitrary fields. In *Proc. 40th FOCS*, pp. 45–50. IEEE Comp. Soc. Press, 1999. [doi:10.1109/SFFCS.1999.814576] 286

[3] PETER BÜRGISSER, MICHAEL CLAUSEN, AND MOHAMMAD AMIN SHOKROLLAHI: *Algebraic Complexity Theory*. Volume 315 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1997. [doi:10.1007/978-3-662-03338-8] 285, 286, 293

[4] PETER BÜRGISSER AND CHRISTIAN IKENMEYER: Geometric complexity theory and tensor rank. In *Proc. 43rd STOC*, pp. 509–518. ACM Press, 2011. [doi:10.1145/1993636.1993704, arXiv:1011.1350] 286

[5] XI CHEN, NEERAJ KAYAL, AND AVI WIGDERSON: Partial derivatives in arithmetic complexity and beyond. *Foundations and Trends in Theoretical Computer Science*, 6(1-2):1–138, 2011. [doi:10.1561/0400000043] 287

[6] WILLIAM FULTON AND JOE HARRIS: *Representation Theory*. Volume 129 of *Graduate Texts in Mathematics*. Springer, 1991. [doi:10.1007/978-1-4612-0979-9] 294

[7] ANKIT GUPTA, PRITISH KAMATH, NEERAJ KAYAL, AND RAMPRASAD SAPTHARISHI: Approaching the chasm at depth four. *J. ACM*, 61(6):33:1–33:16, 2014. Preliminary versions in ECCC and CCC'13. [doi:10.1145/2629541] 287

[8] JONATHAN D. HAUENSTEIN, CHRISTIAN IKENMEYER, AND JOSEPH M. LANDSBERG: Equations for lower bounds on border rank. *Experimental Mathematics*, 22(4):372–383, 2013. [doi:10.1080/10586458.2013.825892] 286

[9] NEERAJ KAYAL: An exponential lower bound for the sum of powers of bounded degree polynomials. *Electron. Colloq. on Comput. Complexity (ECCC)*, TR12-081, 2012. 287

[10] JOSEPH M. LANDSBERG: The border rank of the multiplication of 2×2 matrices is seven. *J. Amer. Math. Soc.*, 19(2):447–459, 2006. [doi:10.1090/S0894-0347-05-00506-0, arXiv:math/0407224] 286

[11] JOSEPH M. LANDSBERG: *Tensors: Geometry and Applications*. Volume 128 of *Graduate Studies in Mathematics*. Amer. Math. Soc., Providence, 2012. [doi:10.1090/gsm/128] 286, 288, 290, 293, 294, 295

[12] JOSEPH M. LANDSBERG: New lower bounds for the rank of matrix multiplication. *SIAM J. Comput.*, 43(1):144–149, 2014. [doi:10.1137/120880276, arXiv:1206.1530] 286

[13] JOSEPH M. LANDSBERG: Nontriviality of equations and explicit tensors in $\mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$ of border rank at least $2m-2$. *J. Pure Appl. Algebra*, 219(8):3677–3684, 2015. [doi:10.1016/j.jpaa.2014.12.016] 289

[14] JOSEPH M. LANDSBERG AND GIORGIO OTTAVIANI: Equations for secant varieties of Veronese and other varieties. *Annali di Matematica Pura ed Applicata*, 192(4):569–606, 2013. [doi:10.1007/s10231-011-0238-6] 287, 289, 290

[15] FRANÇOIS LE GALL: Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th Int. Symp. on Symb. and Alg. Comput. (ISSAC'14)*, pp. 296–303. ACM, New York, 2014. [doi:10.1145/2608628.2608664, arXiv:1401.7714] 286

[16] THOMAS LICKTEIG: A note on border rank. *Inf. Process. Lett.*, 18(3):173–178, 1984. [doi:10.1016/0020-0190(84)90023-1] 286

[17] ALEX MASSARENTI AND EMANUELE RAVIOLO: The rank of $n \times n$ matrix multiplication is at least $3n^2 - 2\sqrt{2}n^{\frac{3}{2}} - 3n$. *Linear Algebra Appl.*, 438(11):4500–4509, 2013. [doi:10.1016/j.laa.2013.01.031] 286

[18] ALEX MASSARENTI AND EMANUELE RAVIOLO: Corrigendum to "The rank of $n \times n$ matrix multiplication is at least $3n^2 - 2\sqrt{2}n^{\frac{3}{2}} - 3n$". *Linear Algebra Appl.*, 445:369–371, 2014. [doi:10.1016/j.laa.2013.12.009] 286

[19] KETAN D. MULMULEY AND MILIND A. SOHONI: Geometric complexity theory I: an approach to the P vs. NP and related problems. *SIAM J. Comput.*, 31(2):496–526, 2001. [doi:10.1137/S009753970038715X] 287

[20] GIORGIO OTTAVIANI: Symplectic bundles on the plane, secant varieties and Lüroth quartics revisited. In *Vector bundles and low codimensional subvarieties: state of the art and recent developments*, volume 21 of *Quad. Mat.*, pp. 315–352. Dept. Math., Seconda Univ. Napoli, Caserta, 2007. [arXiv:math/0702151] 288

[21] CLAUDIO PROCESI: *Lie Groups: An Approach through Invariants and Representations*. Universitext. Springer, 2007. [doi:10.1007/978-0-387-28929-8] 294

[22] ALEXEY V. SMIRNOV: The bilinear complexity and practical algorithms for matrix multiplication. *Comput. Math. Math. Phys.*, 53(12):1781–1795, 2013. [doi:10.1134/S0965542513120129] 286

[23] VOLKER STRASSEN: Gaussian elimination is not optimal. *Numer. Math.*, 13(4):354–356, 1969. [doi:10.1007/BF02165411] 286

[24] VOLKER STRASSEN: Rank and optimal computation of generic tensors. *Linear Algebra Appl.*, 52-53:645–685, 1983. [doi:10.1016/0024-3795(83)80041-X] 286, 288

[25] AVI WIGDERSON: P, NP and mathematics—a computational complexity perspective. In *International Congress of Mathematicians. Vol. I*, pp. 665–712. Eur. Math. Soc., Zürich, 2007. [doi:10.4171/022-1/25] 287

AUTHORS

Joseph M. Landsberg
Professor of Mathematics
Texas A&M University, College Station, TX, USA
jml@math.tamu.edu
http://www.math.tamu.edu/~jml

Giorgio Ottaviani
Professore Ordinario di Geometria
Università di Firenze, Firenze Italy
ottaviani@math.unifi.it
http://web.math.unifi.it/users/ottaviani/

ABOUT THE AUTHORS

JOSEPH M. LANDSBERG works on questions at the interface of algebraic geometry, differential geometry and representation theory. His current research is focused on geometric problems originating in complexity theory. Landsberg obtained his Ph. D. in 1990 from Duke University under the direction of Robert Bryant on minimal submanifolds and a generalization of calibrations. He has directed six Ph. D. theses, is the author of *Tensors: Geometry and Applications*, is a co-author of *Cartan for Beginners*, and has published over fifty research articles. Fall 2014 he co-organized *Algorithms and Complexity in Algebraic Geometry* at the Simons Institute for the Theory of Computing, UC Berkeley, where he was also the Chancellor's Professor and gave a course on geometry and complexity theory whose class notes will hopefully soon be a book.

GIORGIO OTTAVIANI is a Professor of Geometry at the University of Florence, Italy. He received his Ph. D. in mathematics in 1987 at the University of Florence, under the guidance of Vincenzo Ancona, discussing the thesis "Vector bundles on Grassmannians and Quadrics." His M. Sc. advisor was Francesco Gherardelli. He moved to the University of Rome Tor Vergata as an assistant researcher and then to the University of L'Aquila as a professor. Ottaviani was awarded the Foundation Severi Prize in 1989. His first works discussed instanton vector bundles, homogeneous vector bundles, small codimension projective subvarieties, and enjoying the help of computational tools. In the past eight years, Ottaviani has become interested in the field of tensors and their applications, focusing on tensor rank.