

Subrank and Optimal Reduction of Scalar Multiplications to Generic Tensors

Harm Derksen 

Northeastern University, Boston, MA, USA

Visu Makam 

Radix Trading Europe B.V., Amsterdam, The Netherlands

Jeroen Zuiddam 

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands

Abstract

Since the seminal works of Strassen and Valiant it has been a central theme in algebraic complexity theory to understand the *relative* complexity of algebraic problems, that is, to understand which algebraic problems (be it bilinear maps like matrix multiplication in Strassen’s work, or the determinant and permanent polynomials in Valiant’s) can be reduced to each other (under the appropriate notion of reduction).

In this paper we work in the setting of bilinear maps and with the usual notion of reduction that allows applying linear maps to the inputs and output of a bilinear map in order to compute another bilinear map. As our main result we determine precisely how many independent scalar multiplications can be reduced to a given bilinear map (this number is called the *subrank*, and extends the concept of matrix diagonalization to tensors), for essentially all (i.e. generic) bilinear maps. Namely, we prove for a generic bilinear map $T : V \times V \rightarrow V$ where $\dim(V) = n$ that $\theta(\sqrt{n})$ independent scalar multiplications can be reduced to T . Our result significantly improves on the previous upper bound from the work of Strassen (1991) and Bürgisser (1990) which was $n^{2/3+o(1)}$. Our result is very precise and tight up to an additive constant. Our full result is much more general and applies not only to bilinear maps and 3-tensors but also to k -tensors, for which we find that the generic subrank is $\theta(n^{1/(k-1)})$. Moreover, as an application we prove that the subrank is not additive under the direct sum.

The subrank plays a central role in several areas of complexity theory (matrix multiplication algorithms, barrier results) and combinatorics (e.g., the cap set problem and sunflower problem). As a consequence of our result we obtain several large separations between the subrank and tensor methods that have received much interest recently, notably the slice rank (Tao, 2016), analytic rank (Gowers–Wolf, 2011; Lovett, 2018; Bhurshundi–Harsha–Hatami–Kopparty–Kumar, 2020), geometric rank (Kopparty–Moshkovitz–Zuiddam, 2020), and G-stable rank (Derksen, 2020).

Our proofs of the lower bounds rely on a new technical result about an optimal decomposition of tensor space into structured subspaces, which we think may be of independent interest.

2012 ACM Subject Classification Theory of computation \rightarrow Algebraic complexity theory

Keywords and phrases tensors, bilinear maps, complexity, subrank, diagonalization, generic tensors, random tensors, reduction, slice rank

Digital Object Identifier 10.4230/LIPIcs.CCC.2022.9

Related Version *Full Version:* <https://arxiv.org/abs/2205.15168>

Funding *Harm Derksen:* NSF grants IIS-1837985 and DMS-2001460.

Visu Makam: NSF grants CCF-1900460 and the University of Melbourne.

Jeroen Zuiddam: Simons Junior Fellowship and NWO Veni grant VI.Veni.212.284.

Acknowledgements The authors thank Swastik Kopparty for helpful discussions, and JM Landsberg and Jan Draisma for comments.



© Harm Derksen, Visu Makam, and Jeroen Zuiddam;
licensed under Creative Commons License CC-BY 4.0
37th Computational Complexity Conference (CCC 2022).

Editor: Shachar Lovett; Article No. 9; pp. 9:1–9:23

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



1 Introduction

We solve a fundamental problem in algebraic complexity theory about a notion of complexity on bilinear maps (tensors) called the *subrank*, which was introduced by Strassen in [39] in the study of fast matrix multiplication algorithms, and which later found close connections to several hypergraph independence and Ramsey-type problems in combinatorics and tensor methods in these areas (e.g., analytic rank [28, 3] and related notions). Our results improve significantly on previous bounds from the work of Bürgisser [9] and Strassen [41].

In high-level terms, the subrank of a bilinear map is the largest number of independent scalar multiplications that can be reduced to (i.e. “embedded in”) the bilinear map, under the natural algebraic complexity notion of reduction (which we elaborate on in a moment). This definition extends the usual rank of matrices and moreover naturally extends further to multilinear maps (k -tensors). In this paper, we:

- determine the subrank for almost all bilinear maps (i.e. generic bilinear maps),
- prove precise bounds that are accurate up to a small additive constant,
- extend the above results from bilinear maps (3-tensors) also to multilinear maps (k -tensors),
- prove as a technical result (used in the proofs of the above) an optimal decomposition theorem for tensor subspace, decomposing tensor space into very structured subspaces,
- prove, as an application of our upper bound on generic subrank, that the subrank is not additive under the direct sum.

Let us briefly state our asymptotic results here. (We will return to these in more detail in Subsection 1.2.) We prove for any vector space¹ V with $\dim(V) = n$ that almost all bilinear maps $T : V \times V \rightarrow V$ have subrank equal to $\theta(\sqrt{n})$. That is, $\theta(\sqrt{n})$ independent scalar multiplications can be reduced to T . For k -tensors we find the subrank to be $\theta(n^{1/(k-1)})$ on almost all k -tensors.

To prove our results we use methods from algebraic geometry as well as linear algebraic arguments. Our upper bound proof relies on an efficient parametrization of the set of tensors with subrank larger than a given number (as a collection of orbits) and a good dimension estimate of this set. Our lower bound proof relies on arguments involving differentials. At the core we prove a technical result about a very structured decomposition of tensor space which we think may be of independent interest. The main goal here (in the simplest case) is to write tensor space as a sum of tensor subspaces as efficiently as possible (meaning with smallest as possible sum of dimensions) such that each subspace has the special form of a matrix subspace tensored with n -space. Our technical result is that we can do this optimally for any order of tensor space. We discuss the proof methods further in Subsection 1.3.

To get some intuition for the subrank it is instructive to ask how the subrank relates to the better known complexity notion *tensor rank*. Subrank and rank are indeed closely linked, and in a sense they are dual to each other. Indeed, whereas the subrank of a bilinear map T is the maximal number of independent scalar multiplications that can be reduced to T , the rank of T is (among several equivalent definitions) the minimal number of independent scalar multiplications that T reduces to. In this way, the rank measures the computational “cost” of T in terms of scalar multiplications while the subrank measures the “value” of T in terms of independent scalar multiplications.

¹ We will require the base field of the vector space to have the mild property of being algebraically closed (but it will be clear from our techniques that this assumption can be weakened considerably).

1.1 Subrank and generic subrank

Let us now discuss more precisely several equivalent formulations and the meaning of the definition of subrank, and the sense in which we determine the subrank for almost all tensors (the generic subrank).

Subrank of bilinear and multilinear maps

We first stay in the realm of bilinear maps, and will define subrank via the notion of a reduction between bilinear maps, which is really a reduction in the sense of computational complexity. Given two bilinear maps $S : V_1 \times V_2 \rightarrow V_3$ and $T : W_1 \times W_2 \rightarrow W_3$ we say S reduces to T and write $S \leq T$ if there are linear maps $L_1 : V_1 \rightarrow W_1$, $L_2 : V_2 \rightarrow W_2$ and $L_3 : W_3 \rightarrow V_3$ such that $S(v_1, v_2) = L_3(T(L_1(v_1), L_2(v_2)))$ for all $v_1 \in V_1, v_2 \in V_2$. (In the literature, this relation \leq on tensors is often called the *restriction preorder* and when $S \leq T$ one says that “ S is a restriction of T ” [39, 8].) In other words, if $S \leq T$, then any algorithm for T also gives an algorithm for S (with only small computational overhead, namely applying the linear maps L_i to inputs and output). Next, an important basic bilinear map that we use to define subrank is (for any positive integer r) the bilinear map I_r that computes r independent scalar multiplications with scalars from some² field K :

$$I_r : K^r \times K^r \rightarrow K^r : (v_1, v_2) \mapsto ((v_1)_1(v_2)_1, \dots, (v_1)_r(v_2)_r).$$

With these notions set up, the subrank $Q(T)$ of a bilinear map $T : V_1 \times V_2 \rightarrow V_3$ is defined as the largest number r such that I_r reduces to T , that is, $I_r \leq T$. On the other hand, the tensor rank of T is the smallest number r such that $T \leq I_r$. Thus, the subrank $Q(T)$ is indeed the largest number of independent scalar multiplications that can be reduced to T , while the rank is the smallest number of independent scalar multiplications that T reduces to. The above definition of subrank extends directly to multilinear maps $V_1 \times \dots \times V_{k-1} \rightarrow V_k$ by naturally extending the reduction and defining I_r as the multilinear map that computes r independent $(k-1)$ -wise products.

Subrank of tensors.

Alternatively we may phrase the definition of the subrank in the language of tensors, as bilinear maps (and multilinear maps) naturally correspond to these. Let K^{n_1, n_2, n_3} be the space of 3-tensors (3-dimensional arrays) $T = (T_{i,j,k})_{i,j,k}$ with $i \in [n_1], j \in [n_2], k \in [n_3]$. For tensors $S \in K^{n_1, n_2, n_3}$ and $T \in K^{m_1, m_2, m_3}$ we write $S \leq T$ if there are matrices $A \in \text{Mat}_{n_1, m_1}$, $B \in \text{Mat}_{n_2, m_2}$, $C \in \text{Mat}_{n_3, m_3}$ such that S is obtained from T by applying A, B, C to the slices of T , in the sense that

$$S_{i,j,k} = \sum_{a,b,c} A_{i,a} B_{j,b} C_{k,c} T_{a,b,c}$$

for all $i \in [n_1], j \in [n_2], k \in [n_3]$. Let $I_r \in K^{r,r,r}$ be the tensor for which the diagonal entries $(I_r)_{i,i,i}$ are 1 and the other entries are 0. In this language the subrank $Q(T)$ is again the largest number r such that $I_r \leq T$. This definition of subrank also naturally extends to higher-order tensors K^{n_1, \dots, n_k} .

² We will put some mild restrictions on the field in some parts of the paper.

Subrank on “almost all” tensors; generic subrank

It follows from a short argument (which we give later; Proposition 9) that for vector spaces V_1, V_2, V_3 , there is a subset U of all bilinear maps $V_1 \times V_2 \rightarrow V_3$ that is nonempty and Zariski-open (the complement is the zero-set of a finite collection of polynomials) and with constant subrank. This set U thus contains almost all bilinear maps $V_1 \times V_2 \rightarrow V_3$. The subrank on U is called the *generic subrank*. We denote the generic subrank of bilinear maps $K^{n_1} \times K^{n_2} \rightarrow K^{n_3}$ (or equivalently of tensors in K^{n_1, n_2, n_3}) by $Q(n_1, n_2, n_3)$ (and by $Q(n_1, \dots, n_k)$ generally for higher-order tensors).

1.2 Our Results

As our main results we determine (1) the subrank of generic tensors of order three, and, more generally, (2) the subrank of generic tensors of any order $k \geq 3$. Moreover, as a core technical ingredient in our proof of (1) and (2), we prove (3) an optimal decomposition of tensor space into highly structured subspaces, which we think is of independent interest and which may have further applications in algebraic complexity theory. We will now describe each of these results in detail.

(1) The generic subrank of tensors of order three.

We will begin by discussing our results for tensors of order three (or equivalently, for trilinear maps or bilinear maps on vector spaces of appropriate dimensions). Recall that $Q(n) = Q(n, n, n)$ denotes the generic subrank of tensors in $K^{n, n, n}$. In other words, $Q(n)$ is the value that the subrank takes on “essentially all” tensors, or on “randomly chosen” tensors with probability 1. We prove that $Q(n)$ grows as \sqrt{n} .

► **Theorem 1.** *We have $Q(n) = \theta(\sqrt{n})$.*

We thus solve the problem of determining the generic subrank of tensors of order three (and also, with more work, of order k in general as we will discuss below). In computational terms, Theorem 1 states that for “essentially all” bilinear maps $T : K^n \times K^n \rightarrow K^n$ we can reduce the problem of multiplying \sqrt{n} independent pairs of numbers to T , and that this is optimal (under the usual notion of reduction from algebraic complexity theory in which linear maps are applied to the inputs and output of T).

In particular, with Theorem 1 we significantly improve on the previous best upper bound on $Q(n)$ of Strassen and Bürgisser (obtained via the method of “lower support functionals”) which was $Q(n) \leq n^{2/3+o(1)}$.

As an application of Theorem 1 we prove that the subrank is not additive under the direct sum. Namely, we prove that there are bilinear maps $S, T : K^n \times K^n \rightarrow K^n$ such that $Q(T) + Q(S) = \theta(\sqrt{n})$ while $Q(T \oplus S) \geq n$. In other words, it is sometimes possible to reduce many more independent scalar multiplications to a direct sum of bilinear maps than to the bilinear maps separately.³

³ This result is the subrank analogue of the recent result of Shitov [36] that disproved Strassen’s additivity conjecture for tensor rank [37]. For this he constructed a complicated example of 3-tensors S, T (over any infinite field) such that $R(S \oplus T) < R(S) + R(T)$. We discuss additivity results further in Subsection 1.4.

As a direct consequence of Theorem 1, we separate the generic subrank $Q(n)$ from the generic asymptotic subrank $\underline{Q}(n)$ ⁴, for which Strassen proved that the generic value satisfies $\underline{Q}(n) \geq n^{2/3}$ [40, Prop. 3.6]. Moreover, our result gives a large separation between the subrank on the one hand, and the slice rank (and partition rank, for higher order tensors), geometric rank, and G-stable rank on the other hand (whose generic value is full).

Let us now discuss the bound of Theorem 1 in more detail, and in particular get the precise constants right, after which we discuss the extension to higher-order tensors. We (naturally) obtain Theorem 1 in two parts, namely by first proving the following upper bound:

► **Theorem 2.** *We have $Q(n) \leq \lfloor \sqrt{3n-2} \rfloor$.*

And then proving the following essentially matching lower bound:

► **Theorem 3.** *We have $Q(n) \geq 3(\lfloor \sqrt{n/3 + 1/4} - 1/2 \rfloor)$.*

It is not difficult to see that our upper and lower bounds on the generic subrank are very precise. Namely, the difference between the upper bound $\lfloor \sqrt{3n-2} \rfloor$ and the lower bound $3(\lfloor \sqrt{n/3 + 1/4} - 1/2 \rfloor)$ is at most a small additive constant.⁵

We conjecture that our upper bound in Theorem 2 is tight. Through a more sophisticated analysis (which requires as a final component a computer verification that a determinant is nonzero) we prove that for all $1 \leq n \leq 100$ our upper bound is in fact tight, that is: for all $1 \leq n \leq 100$ we have $Q(n) = \lfloor \sqrt{3n-2} \rfloor$.

(2) The generic subrank of k -tensors.

So far we have discussed only tensors of order three. With more elaborate methods we are able to completely extend our above results from tensors of order three to tensors of order k for every $k \geq 3$. Denoting the subrank of a generic tensor in $K^{n, \dots, n}$ of order k by $Q(n, \dots, n)$, we find that $Q(n, \dots, n)$ grows as $n^{1/(k-1)}$.

► **Theorem 4.** *We have $Q(n, \dots, n) = \theta(n^{1/(k-1)})$.*

Again this result directly leads to a large separation between the subrank and the asymptotic subrank (for which the generic value, extending a construction of Strassen, satisfies $\underline{Q}(n) \geq n^{2/k}$), and a separation between the subrank and the slice rank, partition rank, geometric rank and G-stable rank (all of whose generic value is n).

Regarding the precise bounds, we prove Theorem 4 in two parts again. In the first we extend the upper bound of Theorem 2 to order k , in a fully general manner as an upper bound on the generic subrank $Q(n_1, \dots, n_k)$ where the n_i need not be equal, as follows.

► **Theorem 5.** *We have $Q(n_1, \dots, n_k) \leq (\sum_{i=1}^k n_i - (k-1))^{1/(k-1)}$.*

Then finally we extend Theorem 3 to order k tensors, which leads to the asymptotically matching lower bound (which for conciseness we will not write down here more explicitly).

► **Theorem 6.** *We have $Q(n, \dots, n) \geq \Omega(n^{1/(k-1)})$.*

Our proof of this last theorem (and also Theorem 3) makes crucial use of the technical results on tensor space that we discuss in the next section.

⁴ The asymptotic subrank is defined as $\underline{Q}(T) = \lim_{n \rightarrow \infty} Q(T^{\otimes n})^{1/n}$ and $\underline{Q}(n)$ denotes the value of $\underline{Q}(T)$ for a generic tensor $T \in K^{n, n, n}$.

⁵ Indeed, a straightforward direct computation shows that $\lfloor \sqrt{3n-2} \rfloor - 3(\lfloor \sqrt{n/3 + 1/4} - 1/2 \rfloor) \leq 5$.

(3) Technical result: structured subspace decomposition of tensor space

The proofs of our lower bounds (Theorem 3 and Theorem 6) rely on a technical result about a very structured decomposition of tensor space which we think may be of independent interest, so we will describe it here. The main goal here (in the simplest case) is to write tensor space $K^{n,n,n}$ as a sum of tensor subspaces as efficiently as possible (meaning with smallest as possible sum of dimensions) such that each subspace has the form of an $n \times n$ matrix subspace tensored with K^n (so that it becomes a subspace of $K^{n,n,n}$) with the tensoring being applied in any of the three possible directions. Our result is that we can do this optimally for any order of tensor space.

We will begin our discussion with the simplest version, which is for the tensor space $K^{3,3,3}$, as it is the easiest to explain (and in fact also not hard to prove) and forms the basis for the proof of the (harder to prove) general version for the tensor space $K^{n,\dots,n} = (K^n)^{\otimes n}$ of tensors of order n and dimension n in each direction (in our application to the generic subrank we use a blow-up argument so that we can deal with tensors of order k and dimension n in each direction).

Recall that K^{n_1,n_2,n_3} denotes the space of $n_1 \times n_2 \times n_3$ tensors with coefficients in the field K , which we require to be large enough in this part. Let $\text{Mat}_{n,m}$ denote the space of $n \times m$ matrices with coefficients in K . We use a special notation to denote a certain construction of a tensor subspaces given a matrix subspace. Namely, for any given $n_1 \times n_2$ matrix subspace $\mathcal{W} \subseteq \text{Mat}_{n_1,n_2}$ we denote by $\mathcal{W}[3]$ the tensor subspace $\mathcal{W} \otimes K^{n_3} \subseteq K^{n_1,n_2,n_3}$. Analogously, for any matrix subspace $\mathcal{W} \subseteq \text{Mat}_{n_2,n_3}$ we denote by $\mathcal{W}[1] \subseteq K^{n_1,n_2,n_3}$ the tensor subspace obtained by appropriately tensoring \mathcal{W} with K^{n_1} , and for any matrix subspace $\mathcal{W} \subseteq \text{Mat}_{n_1,n_3}$ we denote by $\mathcal{W}[2] \subseteq K^{n_1,n_2,n_3}$ the tensor subspace obtained by appropriately tensoring \mathcal{W} with K^{n_2} .

For the tensor space $K^{3,3,3}$ we have the following optimal decomposition theorem.

► **Theorem 7.** *There exist subspaces $\mathcal{X}_i \subseteq \text{Mat}_{3,3} = K^3 \otimes K^3$, each of dimension 3, such that*

$$K^{3,3,3} = \mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3].^6$$

Comparing dimensions, we have $\dim K^{3,3,3} = 3 \cdot 3 \cdot 3 = \dim \mathcal{X}_1[1] + \dim \mathcal{X}_2[2] + \dim \mathcal{X}_3[3]$ and so the decomposition in Theorem 7 is optimal. In other words, it is a direct sum decomposition of $K^{3,3,3}$.

The requirement in Theorem 7 that the \mathcal{X}_i all have dimension 3 is crucial to make the theorem interesting, as without this requirement we could “decompose” $K^{3,3,3}$ simply as $K^{3,3,3} = \mathcal{X}_1[1]$ with $\mathcal{X}_1 = \text{Mat}_{3,3}$. Interestingly, the analogous statement of Theorem 7 for any matrix space $K^{n,n}$ is false. That is, for every positive integer n there are no subspaces $\mathcal{X}_i \subseteq K^n$, each of dimension $n/2$, such that $K^{n,n} = \mathcal{X}_1[1] + \mathcal{X}_2[2]$. This is saying that if we pick a row space R and a column space C of dimension $n/2$, then it is not possible to write every $n \times n$ matrix A as $A_1 + A_2$ where the row space of A_1 is in R and the column space of A_2 is in C . On the other hand, 3-tensors do not suffer from this malady: we can write any tensor as a sum of three tensors, whose “row_{*i*}”-space are asked to be in a generic space of the right dimension. (Here we can think of $W[i]$ as tensors whose “row_{*i*}”-space is W .) Therefore, Theorem 7 is demonstrating an interesting phenomenon that occurs in tensor space but not in matrix space.

⁶ We note that the existence of subspaces \mathcal{X}_i with this property is equivalent to generic subspaces \mathcal{X}_i having this property (i.e. there being a non-empty Zariski-open set of triples \mathcal{X}_i with this property).

Theorem 7 (as opposed to the upcoming generalization to $(K^n)^{\otimes n}$) is not difficult to prove. Indeed, one may choose the matrix subspaces \mathcal{X}_i randomly and then for such an explicit choice verify directly that they satisfy the claim (and this approach will work with high probability). In fact, there are even valid choices of the \mathcal{X}_i such that each \mathcal{X}_i is spanned by a matrix with coefficients in $\{0, 1\}$.

Next we discuss the higher-dimensional version of Theorem 7 in which $K^{3,3,3}$ is generalized to n -tensor space $(K^n)^{\otimes n}$. We naturally extend our previous notation so that for every tensor subspace $\mathcal{W} \subseteq (K^n)^{\otimes(n-1)}$ we define, for any $i \in \{1, \dots, n\}$, the tensor subspace $\mathcal{W}[i] \subseteq (K^n)^{\otimes n}$ by tensoring \mathcal{W} with K^n in one of the n possible ways.

By a recursive construction with $K^{3,3,3}$ as a base case, we find the following optimal decomposition of n -tensor space $(K^n)^{\otimes n}$ for all $n \geq 3$.

► **Theorem 8.** *For every integer $n \geq 3$ there exist subspaces $\mathcal{X}_i \subseteq (K^n)^{\otimes(n-1)}$ of dimension n^{n-2} such that*

$$(K^n)^{\otimes n} = \mathcal{X}_1[1] + \mathcal{X}_2[2] + \dots + \mathcal{X}_n[n].$$

Again, since $\dim K^{n, \dots, n} = n^n = n \cdot n \cdot n^{n-2} = \sum_{i=1}^n \dim \mathcal{X}_i[i]$, the decomposition in Theorem 8 is optimal in terms of dimension and hence a direct sum decomposition of $(K^n)^{\otimes n}$.

Theorem 8 is the theorem we use to prove the general generic subrank lower bound Theorem 6. However, the methods we introduce in the process of proving Theorem 8 allow us much more generally for other choices of positive integers n_1, \dots, n_k to construct optimal decompositions $K^{n_1, \dots, n_k} = \mathcal{X}_1[1] + \mathcal{X}_2[2] + \dots + \mathcal{X}_k[k]$ from known decompositions. This leads to a natural fundamental mathematical question (with potentially other applications) of what choices of n_1, \dots, n_k and $\dim \mathcal{X}_i$ allow such decompositions.

1.3 Technical Overview

We give a brief technical overview of the methods and ideas that we use in our proofs.

Upper bounds on generic subrank

The high-level approach in our proof of the upper bound on the generic subrank in Theorem 2 (and similarly for the general case of Theorem 5) is as follows. For any nonnegative integer we consider the set \mathcal{C}_r of tensors in $K^{n,n,n}$ with subrank at least r . We argue that the generic subrank is precisely the largest r such that the dimension of \mathcal{C}_r equals the dimension n^3 of the full space $K^{n,n,n}$. We then prove the core ingredient, namely the dimension upper bound $\dim(\mathcal{C}_r) \leq n^3 - r(r^2 - 3n + 2)$. This information leads to the desired result, since if we let t be the generic subrank $Q(n)$, then we must by the above have $n^3 = \dim(\mathcal{C}_t) \leq n^3 - t(t^2 - 3n + 2)$, from which we directly deduce that $t \leq \sqrt{3n - 2}$, that is, we obtain the bound $Q(n) \leq \sqrt{3n - 2}$ of Theorem 2.

To prove the aforementioned dimension upper bound on \mathcal{C}_r that is the core ingredient in the above argument we employ the idea of providing a (non-injective) parametrization of \mathcal{C}_r , compute the dimension of the parameter space, and then subtracting the dimension “over-count” (the fiber dimension under the parametrization). For this we first define a set X_r of tensors in $K^{n,n,n}$ of a special form, namely whose $[r] \times [r] \times [r]$ subtensor is zero except for the diagonal which is nonzero. Then the elements of X_r clearly have subrank at least r (and are thus in \mathcal{C}_r). The important point is that, by applying all possible basis transformations to the tensors in X_r we obtain all of \mathcal{C}_r . Thus X_r together with the group of all basis transformations provide the parametrization of \mathcal{C}_r . Technically, we describe

this by saying that the map $\psi_r : \mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r \rightarrow K^{n,n,n}$ that maps (A, B, C, T) to $(A \otimes B \otimes C)T$ has image precisely \mathcal{C}_r . Now the computation to upper bound $\dim(\mathcal{C}_r)$ consists of computing the dimension of the domain $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r$ and subtracting the dimension of a general fiber of ψ_r , which we carry out to arrive at the dimension upper bound stated earlier.

Lower bounds on generic subrank

The high-level approach in our proof of the lower bound on the generic subrank in Theorem 3 (and the general Theorem 6) is as follows. We use notation defined in the upper bound proof discussion and the results section (Subsection 1.2). Our proof reduces the problem of lower bounding the generic subrank $Q(n_1, n_2, n_3)$ to a problem of constructing tensor space decompositions of a specific form (which we discuss further in the next section). Namely we prove that, for $r \leq n_1, n_2, n_3$ (with a technical condition), if $\mathcal{X}_i \subseteq \mathrm{Mat}_{r,r}$ are subspaces of dimension $n_i - r$ for $i = 1, 2, 3$ such that

$$\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{r,r,r},^7$$

then $Q(n_1, n_2, n_3) \geq r$. (And we prove the analogous statement for higher-order tensors.) Finding such \mathcal{X}_i we discuss in the next section. The proof of the lower bound given the \mathcal{X}_i goes as follows.

Recall the map $\psi_r : \mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r \rightarrow K^{n,n,n}$ whose image we already claimed is the set \mathcal{C}_r of tensors of subrank at least r . To reach our goal we want to find conditions that imply that the image of ψ_r has full dimension n^3 and thus is Zariski-dense in $K^{n,n,n}$. To do this we use the notion of the differential $d\psi_r$ of ψ_r and a general method that says that the dimension of a map can be computed as the rank of the differential at a “generic point”.

The differential $d\psi_r$ at the point (g_1, g_2, g_3, T) is the map

$$(d\psi_r)_{(g_1, g_2, g_3, T)} : \mathrm{Mat}_{n,n} \times \mathrm{Mat}_{n,n} \times \mathrm{Mat}_{n,n} \times Y_r \rightarrow K^{n,n,n}$$

where Y_r is the tangent space of X_r , given by

$$(A, B, C, S) \mapsto ((A \otimes g_2 \otimes g_3) + (g_1 \otimes B \otimes g_3) + (g_1 \otimes g_2 \otimes C))T + (g_1 \otimes g_2 \otimes g_3)S.$$

Analyzing this map we compute its image which leads to the aforementioned lower bound statement on the generic subrank. In fact we prove a stronger lower bound, which in characteristic 0 characterizes the generic subrank precisely, but with a harder to analyze condition, in Theorem 21. In particular, if the characteristic of K is 0, then Theorem 21 says that $Q(n_1, n_2, n_3)$ is given by the smallest number r such that $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r}$ for generic subspaces $\mathcal{X}_i \subseteq \mathrm{Mat}_{r,r}$ of dimension $n_i - r$, where $W_r \subseteq K^{r,r,r}$ is the subspace of tensors such that $T_{ijk} = 0$ if i, j, k are all different. (In other characteristics this number r gives a lower bound.)

Constructions of tensor space decompositions

To prove Theorem 8 we introduce general methods to construct the optimal tensor space decompositions as described in the theorem from existing ones. We then give some small constructions and combine these in multiple recursions to achieve the required object.

⁷ where $\mathcal{X}_1[1] \subseteq K^{r,r,r}$ denotes \mathcal{X}_1 tensored with K^r in the first tensor leg, $\mathcal{X}_2[2]$ denotes \mathcal{X}_2 tensored with K^r in the second tensor leg, and $\mathcal{X}_3[3]$ denotes \mathcal{X}_3 tensored with K^r in the third tensor leg

To set this up we take a very general approach in which we study direct sum decompositions

$$K^{n_1, n_2, \dots, n_k} = \mathcal{X}_1[1] + \dots + \mathcal{X}_k[k]$$

where $\mathcal{X}_i \subseteq K^{n_1} \otimes \dots \otimes K^{n_{i-1}} \otimes K^{n_{i+1}} \otimes \dots \otimes K^{n_k}$ is a tensor subspace and $\mathcal{X}_i[i]$ denotes the subspace obtained by tensoring \mathcal{X}_i with K^{n_i} as the i th tensor factor. Let a_i be the dimension of \mathcal{X}_i . We are interested in which values of n_1, \dots, n_k and a_1, \dots, a_k allow for a decomposition of the above form. Writing these numbers into a $2 \times k$ matrix

$$\begin{bmatrix} n_1 & n_2 & \dots & n_k \\ a_1 & a_2 & \dots & a_k \end{bmatrix}$$

we let \mathcal{S} be the set of all such matrices for which a decomposition exists satisfying the parameters given in the matrix. Then Theorem 8 corresponds to proving that the $2 \times n$ matrix

$$\begin{bmatrix} n & n & \dots & n \\ n^{n-2} & n^{n-2} & \dots & n^{n-2} \end{bmatrix} \tag{1}$$

is an element of \mathcal{S} .

Next we observe that there are some simple constructions and properties of elements in \mathcal{S} , such as: if a matrix is in \mathcal{S} then if we permute its columns it is still in \mathcal{S} and the matrix $\begin{bmatrix} n \\ 1 \end{bmatrix}$ is in \mathcal{S} . With slightly more work we can give direct constructions for

$$\begin{bmatrix} 2 & 2 & 2 \\ 0 & 3 & 1 \end{bmatrix} \tag{2}$$

being an element of \mathcal{S} , for instance.

In order to construct more elements of \mathcal{S} we prove a ‘‘direct sum construction’’ that given two elements in \mathcal{S} combines them to get a new one. Namely, this result gives that, if

$$\begin{bmatrix} n_1 & n_2 & \dots & n_{k-1} & n'_k \\ a'_1 & a'_2 & \dots & a'_{k-1} & a_k \end{bmatrix} \in \mathcal{S}$$

and

$$\begin{bmatrix} n_1 & n_2 & \dots & n_{k-1} & n''_k \\ a''_1 & a''_2 & \dots & a''_{k-1} & a_k \end{bmatrix} \in \mathcal{S}$$

then we have

$$\begin{bmatrix} n_1 & n_2 & \dots & n_k \\ a_1 & a_2 & \dots & a_k \end{bmatrix} \in \mathcal{S}$$

where $n_k = n'_k + n''_k$, and $a_i = a'_i + a''_i$ for $i = 1, 2, \dots, k - 1$.

Finally, from some simple base cases including (2) we give a construction for

$$\begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$$

being an element of \mathcal{S} and via an elaborate argument with multiple recursions arrive at the matrix in (1) being an element of \mathcal{S} , which is the ingredient required in our proof of Theorem 6. It is natural to ask what precisely are all elements of \mathcal{S} , which we leave as an open problem.

1.4 Related Work

The previous best bound on the generic subrank $Q(n)$ was the upper bound $n^{2/3+o(1)}$ which follows from the work of Bürgisser [9, Satz 2.8]⁸ as part of the broader research program on the theory of asymptotic spectra of tensors (motivated by the study of matrix multiplication algorithms). The proof of this bound relies on the method of “lower support functionals” introduced by Strassen in [41] (see also the more recent surveys on this topic in [10, 44, 43]) and the properties of these that he proves there. This method recovers certain asymptotic information about tensors, which importantly is monotone under the restriction preorder and normalized on diagonal tensors so that it provides an upper bound on the subrank (in a manner that is very different from the approach that we take to prove our optimal upper bound). Bürgisser’s analysis of this method on generic tensors consists of proving that the support of a generic tensor is large for any choice of basis and a combinatorial study of a certain type of covering of these supports, which leads to the aforementioned $n^{2/3+o(1)}$ upper bound.

Recent research has brought about a rich collection of tensor methods that are in a similar “regime” as the subrank (for instance, they are all monotone under restriction), each with their own properties and applications in complexity theory and combinatorics. Notable are the slice rank [42] and closely related partition rank [30], the analytic rank [20, 28, 3], the geometric rank [26, 18] and G-stable rank [14]. Some important applications of these methods include new bounds on cap sets [42, 16], new bounds on the sunflower problem [31], determining the border subrank of matrix multiplication [26], and proving matrix multiplication barriers [6, 1, 11]. Many strong connections have been shown among these parameters. In particular, Derksen [14] showed that the G-stable rank is equal to the slice rank up to a constant factor, and Cohen–Moshkovitz [12, 13] showed (over large fields) that the partition rank, analytic rank and geometric rank are equal up to a constant factor, the culmination of a long line of work on this topic [21, 24, 2, 22, 23, 29]. All of the aforementioned tensor parameters are lower bounded by the subrank.⁹ Our results provide a large separation on almost all tensors between the subrank and all other aforementioned parameters, as the generic subrank satisfies $Q(n) = \theta(\sqrt{n})$ whereas the generic value of all other parameters is the maximal value n .

The study of “generic” or “typical” complexity in algebraic complexity theory goes back to at least Strassen’s paper “Rank and optimal computation of generic tensors” [38], in which he determines the tensor rank of almost all tensors (i.e. generic tensors). His result is that the rank $R(n_1, n_2, n_3)$ of almost all tensors in K^{n_1, n_2, n_3} grows as $n_1 n_2 n_3 / (n_1 + n_2 + n_3 - 2)$ and a description is given of “perfect shapes” (n_1, n_2, n_3) for which precisely equality $R(n_1, n_2, n_3) = n_1 n_2 n_3 / (n_1 + n_2 + n_3 - 2)$ holds. (It is still a fundamental open problem to find explicit tensors in $K^{n, n, n}$ with rank close to n^2 [5] with important implications to formula lower bounds [32].) For the shape $(n, n, 3)$ with n odd this work provides an equation that determines whether a tensor has typical rank or not. These equations lead to several later generalizations [27, 25] and subsequently precise barrier results for “rank methods” [15, 17]. Our result rather than generic rank determines the generic subrank. The rank of a tensor being the smallest number of scalar multiplications that the tensor reduces to, the subrank is a natural dual to the tensor rank. We note that combining the results on generic subrank and generic rank, we

⁸ [9, Satz 2.8] determines the generic value of a tensor parameter called the “lower support functional” (unteren Trägerfunktional) which upper bounds the subrank as proven in [41].

⁹ The analytic rank requires a natural normalization for this to be true, that is, one should normalize it by the analytic rank of a full-dimensional diagonal tensor.

see that reducing scalar multiplications to a generic tensor and reducing the generic tensor back to scalar multiplications necessarily induces a great loss, as $Q(n, n, n) = \theta(\sqrt{n})$ is much smaller than $R(n, n, n) = \theta(n^2)$.

The study of additivity results is a central theme in complexity theory and mathematics. In algebraic complexity, it has been long known (and crucial in the design of matrix multiplication algorithms) that the border rank of tensors (the approximative version of tensor rank) is not additive under the direct sum [35] (see also [4, Lemma 7.1] and [8, (15.12)]). Strassen conjectured the tensor rank to be additive under the direct sum, but this was disproved recently by Shitov [36]. On the other hand, the analytic rank [28], geometric rank [26], G-stable rank [14] and slice rank [19] were shown to be additive recently. The subrank, as we show, is however not additive. Our proof of this relies on writing a tensor as a sum of two generic tensors, which is reminiscent of methods that Razborov [33] uses to prove a linear upper bound on submodular complexity measures of boolean functions (see also [34]).

1.5 Paper Organization

In Section 2 we prove the upper bound theorems Theorem 2 and Theorem 5. In Section 3 we prove the characterization of generic subrank that allows us to reduce the problem of determining its value to the tensor subspace decomposition problem. In Section 4 we solve the tensor subspace decomposition problem thus completing the lower bound proof. In Section 5 we use our upper bounds to prove that the subrank is not additive under direct sum. In Section 6 we discuss several natural related open problems.

2 Upper bounds on generic subrank

In this section, we prove our upper bounds on the generic subrank of tensors. We give a detailed proof in the case of 3-tensors, and then generalize to tensors of all orders.

2.1 Tensors of order three

The techniques we use are familiar in invariant theory and representation theory. The main idea is to take advantage of the fact that subrank is invariant under a large group of symmetries, namely change of basis on each tensor leg. Further, this group of symmetries has excellent algebraic properties which can often be leveraged to remarkable effect.

First and foremost, we have to argue that generic subrank is a valid notion, which we do in the following proposition. The proof is a standard argument which we will discuss because it naturally uses some ingredients that we will use later on.

► **Proposition 9.** *For every n there is a non-empty Zariski-open subset $U \subseteq K^{n,n,n}$ and integer r such that for all $T \in U$ we have $Q(T) = r$.*

For any n , the number r given by Proposition 9 is unique since any two non-empty Zariski-open subsets $U_1, U_2 \subseteq K^{n,n,n}$ must intersect ($K^{n,n,n}$ is irreducible). This number r we call the generic subrank $Q(n)$. Similarly we define the generic subrank $Q(n_1, \dots, n_k)$ of K^{n_1, \dots, n_k} .

We discuss a couple of preparatory results before giving the proof of Proposition 9. We define X_r to be the set of tensors in $K^{n,n,n}$ whose $[r] \times [r] \times [r]$ subtensor is zero except for the diagonal entries in $[r] \times [r] \times [r]$ which are all nonzero,

$$X_r = \{T \in K^{n,n,n} \mid T_{ijk} = 0 \text{ for } (i, j, k) \in [r]^3 \setminus \{(i, i, i) \mid i \in [r]\} \text{ and } T_{i,i,i} \neq 0 \text{ for } i \in [r]\}.$$

9:12 Subrank and Optimal Reduction of Scalar Multiplications to Generic Tensors

We let ψ_r be the map that applies basis transformations to elements of X_r ,

$$\begin{aligned} \psi_r : \mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r &\longrightarrow K^{n,n,n} \\ (A, B, C, T) &\longmapsto (A \otimes B \otimes C)T. \end{aligned}$$

We define \mathcal{C}_r to be the set of tensors in $K^{n,n,n}$ whose subrank is at least r ,

$$\mathcal{C}_r = \{T \in K^{n,n,n} \mid \mathrm{Q}(T) \geq r\}.$$

► **Lemma 10.** *The image of ψ_r is precisely \mathcal{C}_r .*

Proof. To prove $\mathrm{im}(\psi_r) = \mathcal{C}_r$ we show both inclusions.

First we will prove $\mathrm{im}(\psi_r) \subseteq \mathcal{C}_r$. As a first step we will prove that $X_r \subseteq \mathcal{C}_r$. Let $T \in X_r$. Let $T_{i,i,i} = \lambda_i \neq 0$. Let

$$A = \begin{pmatrix} \mathrm{Id}_r & 0 \\ 0 & 0 \end{pmatrix}$$

where Id_r denotes the identity matrix of size $r \times r$ and let

$$B = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where D is a diagonal matrix of size $r \times r$ whose diagonal entries are $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_r^{-1}$. It is easy to check that $(A \otimes A \otimes B) \cdot T = I_r$. Thus $X_r \subseteq \mathcal{C}_r$. Since subrank is invariant under the action of $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n$, we see that \mathcal{C}_r is $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n$ invariant, so we deduce that $\mathrm{im}(\psi_r) = (\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n) \cdot X_r \subseteq \mathcal{C}_r$.

Now we will prove $\mathcal{C}_r \subseteq \mathrm{im}(\psi_r)$. Let $T \in \mathcal{C}_r$. Then, there exist $A, B, C \in \mathrm{Mat}_{r,n}$ such that $(A \otimes B \otimes C) \cdot T = I_r$. Let

$$\tilde{A} = \begin{pmatrix} A \\ * \end{pmatrix}$$

be a completion of A to a full rank $n \times n$ matrix, and similarly define \tilde{B} and \tilde{C} . This is possible because A, B, C must all have rank r . Then $(\tilde{A} \otimes \tilde{B} \otimes \tilde{C}) \cdot T \in X_r$, which implies that $T \in \mathrm{im}(\psi_r)$. Thus $\mathcal{C}_r \subseteq \mathrm{im}(\psi_r)$. ◀

We now give the proof of the existence of the generic subrank (Proposition 9). The proof is standard and may safely be skipped.

Proof of Proposition 9. The map Q attains only finitely many values on $V = K^{n,n,n}$, namely $\{0, 1, \dots, n\}$. Thus it has finitely many fibers $P_i = \mathrm{Q}^{-1}(i) \subseteq V$. Each P_i is a constructible set, since $P_i = \mathcal{C}_i \setminus \mathcal{C}_{i+1}$ and \mathcal{C}_i is constructible as a consequence of Lemma 10. Then $\bigcup_{i=0}^n P_i = V$ and so $\bigcup_{i=0}^n \overline{P}_i = V$. However, V is irreducible so there must be an i such that $\overline{P}_i = V$. Since P_i is constructible it contains a subset $U \subseteq P_i$ that is non-empty and Zariski-open in $\overline{P}_i = V$ (this is a general fact, see e.g. [7]). This U and i satisfy the claim. ◀

Now that we have established that the generic subrank exists, we continue to prove our upper bound on it. The following simple lemma is straightforward, but crucial:

► **Lemma 11.** *The generic subrank $\mathrm{Q}(n) = \max\{r \mid \dim(\mathcal{C}_r) = n^3\}$.*

Proof. As above, let P_i denote the subset of tensors with subrank i . Then, by definition of $\mathrm{Q}(n)$ and the fact that $\mathcal{C}_r = \bigsqcup_{i=r}^n P_i$, we deduce that $\dim(\mathcal{C}_r) = n^3$ if and only if $\mathcal{C}_r \supseteq P_{\mathrm{Q}(n)}$ if and only if $r \leq \mathrm{Q}(n)$. ◀

► **Proposition 12.** *The following is an upper bound for the dimension of \mathcal{C}_r :*

$$\dim(\mathcal{C}_r) \leq 3n^2 + (n^3 - r^3 + r) - 3(n(n - r) + r) = n^3 - r(r^2 - 3n + 2).$$

Proof. Consider the map ψ above. The theorem on dimension of fibres says that

$$\dim(\mathcal{C}_r) = \dim(\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r) - \dim(\text{general fiber of } \psi_r).$$

We see that $\dim(\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n \times X_r) = 3n^2 + (n^3 - r^3 + r)$ because each GL_n contributes n^2 to the dimension and X_r is a Zariski-open subset of a linear subspace of $K^{n,n,n}$ of dimension $(n^3 - r^3 + r)$. Thus, to compute $\dim(\mathcal{C}_r)$, we only need to compute the dimension of the general fiber of ψ_r . Note that the dimension of any fiber is at most the dimension of the general fiber, so it suffices to find a lower bound on the dimension of all fibers.

Suppose $T \in \mathcal{C}_r$, then $T = (g_1 \otimes g_2 \otimes g_3) \cdot S$ for some $S \in X_r$ by Lemma 10. It is easy to see that $\dim(\psi^{-1}(T)) = \dim(\psi^{-1}(S))$ since $\psi^{-1}(T)$ and $\psi^{-1}(S)$ are isomorphic as varieties – indeed this follows from the observation that $(A, B, C, U) \in \psi^{-1}(T) \iff (g_1^{-1}A, g_2^{-1}B, g_3^{-1}C, U) \in \psi^{-1}(S)$. Thus, it suffices to lower bound the dimension of $\psi^{-1}(S)$ for $S \in X_r$.

Let $S \in X_r$. Let $L \subseteq \mathrm{GL}_m$ be the subset of matrices of the form

$$\begin{pmatrix} D & 0 \\ * & * \end{pmatrix}$$

where D is a diagonal matrix of size $r \times r$. It is easy to see that $\dim(L) = n(n - r) + r$. For $A, B, C \in L$, it is easy to see that $(A \otimes B \otimes C) \cdot S \in X_r$. Thus, $(A^{-1}, B^{-1}, C^{-1}, (A \otimes B \otimes C) \cdot S) \in \psi^{-1}(S)$. In particular, this means that $\dim(\psi^{-1}(S)) \geq 3 \dim(L) = 3(n(n - r) + r)$. Thus, we conclude that $\dim(\text{generic fiber}) \geq 3(n(n - r) + r)$ and so

$$\dim(\mathcal{C}_r) \leq 3n^2 + (n^3 - r^3 + r) - 3(n(n - r) + r) = n^3 - r(r^2 - 3n + 2). \quad \blacktriangleleft$$

Now, we have everything necessary to prove the upper bound for the subrank of 3-tensors.

Proof of Theorem 2. Suppose the subrank of a generic tensor in $K^{n,n,n}$ is $t = Q(n)$. Then we know that $n^3 = \dim(\mathcal{C}_t)$ by Lemma 11 and $\dim(\mathcal{C}_t) \leq n^3 - t(t^2 - 3n + 2)$ by Proposition 12. Thus, we must have $t^2 - 3n + 2 \leq 0$, so $t \leq \sqrt{3n - 2}$. Hence, we have $Q(n) \leq \sqrt{3n - 2}$ as desired. \blacktriangleleft

2.2 Higher-order tensors

In the general case, where we look at tensors in K^{n_1, n_2, \dots, n_k} , we define analogously the objects X_r and \mathcal{C}_r , the map $\psi : \prod_{i=1}^k \mathrm{GL}_{n_i} \times X_r \rightarrow \mathcal{C}_r$, etc.

Proof of Theorem 5. We obtain analogously that

$$\dim(\mathcal{C}_r) = \left(\sum_{i=1}^k n_i^2 \right) + \left(\prod_{i=1}^k n_i - r^k + r \right) - \sum_{i=1}^k (n_i(n_i - r) + r) = \prod_{i=1}^k n_i - r \left(r^{k-1} - \sum_{i=1}^k n_i + (k-1) \right).$$

Suppose the subrank of a generic tensor in K^{n_1, n_2, \dots, n_k} is $t = Q(n_1, \dots, n_k)$. Then, we have

$$\prod_{i=1}^k n_i = \dim(\mathcal{C}_t) \leq \prod_{i=1}^k n_i - t \left(t^{k-1} - \sum_{i=1}^k n_i + (k-1) \right),$$

so we get that $t^{k-1} - \sum_{i=1}^k n_i + (k-1) \leq 0$, so that

$$Q(n) = t \leq \left(\sum_{i=1}^k n_i - (k-1) \right)^{\frac{1}{k-1}}$$

as desired. ◀

3 Lower bounds on generic subrank

In this section, we describe the technique we use to show lower bounds on generic subrank. In order to make this technique effective, we will need some explicit constructions of linear subspaces which we postpone to the next section.

First, we introduce some notation. We identify Mat_{n_2, n_3} with $K^{n_2} \otimes K^{n_3}$ in the standard way. For a linear subspace $\mathcal{X} \subseteq \text{Mat}_{n_2, n_3}$, we define a linear subspace

$$\mathcal{X}[1] = K^{n_1} \otimes \mathcal{X} \subseteq K^{n_1, n_2, n_3}.$$

The linear subspace $\mathcal{X}[1]$ consists precisely of the tensors whose slices in the first direction (i.e., matrices $(T_{1jk})_{j,k}, (T_{2jk})_{j,k}, \dots, (T_{njk})_{j,k}$) are in \mathcal{X} . Similarly, we define $\mathcal{X}[2]$ (resp. $\mathcal{X}[3]$) as the set of tensors whose slices in the second (resp. third) direction are in \mathcal{X} .

The main idea behind proving our lower bounds is the following result:

► **Theorem 13.** *Let $r \leq n_1, n_2, n_3$ such that $n_i - r \leq r^2$.¹⁰ Let $\mathcal{X}_i \subseteq \text{Mat}_{r,r}$ be a generic subspace of dimension $n_i - r$ for $i = 1, 2, 3$. Suppose $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{r,r,r}$, then*

$$Q(n_1, n_2, n_3) \geq r.$$

In fact, we can prove a stronger version of the above theorem, which we state as Theorem 21. However, the downside of this stronger version is that it has a hypothesis that is more difficult to work with.

To prove that r is a lower bound for the generic subrank, we need to show that the image of ψ_r has dimension n^3 , that is, the image is Zariski-dense in $K^{n,n,n}$. The dimension of the image of a map can be captured by the rank of the differential at a generic point – an idea that is familiar to differential geometers and algebraic geometers alike. In arbitrary characteristic, we know that the rank of the differential at a generic point is a lower bound for the dimension of the image, which is sufficient for proving lower bounds. In characteristic 0, the image of the rank of the differential at a generic point is equal to the dimension of the image. Consequently, we are able to obtain an exact linear algebraic characterization for generic subrank in Theorem 21.

Recall that we defined the map

$$\psi_r : \text{GL}_{n_1} \times \text{GL}_{n_2} \times \text{GL}_{n_3} \times X_r \rightarrow K^{n_1, n_2, n_3},$$

where

$$X_r = \{T \in K^{n_1, n_2, n_3} \mid T_{ijk} = 0 \text{ for } (i, j, k) \in [r]^3 \setminus \{(i, i, i) \mid i \in [r]\} \text{ and } T_{i,i,i} \neq 0 \text{ for } i \in [r]\}.$$

Observe that

$$Y_r = \{T \in K^{n_1, n_2, n_3} \mid T_{ijk} = 0 \text{ for } (i, j, k) \in [r]^3 \setminus \{(i, i, i) \mid i \in [r]\}\}$$

¹⁰Without this assumption it does not mean anything to have a generic subspace of dimension $n_i - r$. Moreover, if $n_i - r > r^2$, then it is easy to see that $Q(n_1, n_2, n_3) \geq r$.

is the tangent space of X_r . The differential at a point (g_1, g_2, g_3, T) is

$$(d\psi_r)_{(g_1, g_2, g_3, T)} : \text{Mat}_{n_1, n_1} \times \text{Mat}_{n_2, n_2} \times \text{Mat}_{n_3, n_3} \times Y_r \longrightarrow K^{n_1, n_2, n_3}$$

given by

$$(A, B, C, S) \longmapsto ((A \otimes g_2 \otimes g_3) + (g_1 \otimes B \otimes g_3) + (g_1 \otimes g_2 \otimes C))T + (g_1 \otimes g_2 \otimes g_3)S.$$

► **Lemma 14.** *If for generic $(g_1, g_2, g_3, T) \in \text{GL}_{n_1} \times \text{GL}_{n_2} \times \text{GL}_{n_3} \times X_r$ we have*

$$\text{im}((d\psi_r)_{(g_1, g_2, g_3, T)}) = K^{n_1, n_2, n_3},$$

then $Q(n_1, n_2, n_3) \geq r$. The converse holds if the characteristic of K is 0.

Proof. If the rank of the differential $d\psi_r$ at a generic point is full, then the image of ψ_r must be full dimensional, that is, Zariski-dense (and constructible). Every tensor in the image of ψ_r has subrank $\geq r$. Hence, the generic subrank must be at least r .

Assume now that characteristic of K is zero. If the rank of $d\psi_r$ at a generic point is not full, then the image of ψ_r is not full dimensional, that is, the set of tensors having subrank $\geq r$ is not Zariski-dense, so we get $Q(n_1, n_2, n_3) < r$, thereby proving the converse. ◀

We use the following equivariance property of the differential.

► **Lemma 15.** *We have for all $g_i \in \text{GL}_{n_i}$ and $T \in X_r$ that*

$$\text{im}((d\psi_r)_{(g_1, g_2, g_3, T)}) = (g_1 \otimes g_2 \otimes g_3)(\text{im}((d\psi_r)_{(I, I, I, T)})).$$

Proof. We see directly that

$$(d\psi_r)_{(g_1, g_2, g_3, T)}(A, B, C, S) = (g_1 \otimes g_2 \otimes g_3)(d\psi_r)_{(I, I, I, T)}(g_1^{-1}A, g_2^{-1}B, g_3^{-1}C, S)$$

which implies the claim. ◀

► **Corollary 16.** *If for generic $T \in X_r$ we have that $\text{im}((d\psi_r)_{(I, I, I, T)}) = K^{n_1, n_2, n_3}$, then $Q(n_1, n_2, n_3) \geq r$. The converse holds if the characteristic of K is 0.*

So, let us now analyze more carefully the image of $(d\psi_r)_{(I, I, I, T)}$ for a generic tensor T . Below, we write d for $(d\psi_r)_{(I, I, I, T)}$ for notational simplicity. Thus d is given by

$$d(A, B, C, S) = ((A \otimes I \otimes I) + (I \otimes B \otimes I) + (I \otimes I \otimes C)) \cdot T + S.$$

The map d is a linear map, so we see that

$$\text{im } d = \sum_{i=1}^3 d(\text{Mat}_{n_i, n_i}) + d(Y_r)$$

where we use the notation $d(Y_r) = d(0, 0, 0, Y_r)$, $d(\text{Mat}_{n_1, n_1}) = d(\text{Mat}_{n_1, n_1}, 0, 0, 0)$, etc.

► **Lemma 17.** *The image of Y_r under the differential map d is*

$$d(Y_r) = Y_r$$

Proof. Observe that for any $S \in Y_r$, we have $d(0, 0, 0, S) = S$. ◀

Let $L_i = [T_{ijk}]_{j,k}$ for $1 \leq i \leq n_1$. These just split the tensor in the first direction as a stack of n matrices. Let \mathcal{L} denote the span of the L_i s.

► **Lemma 18.** *The image of Mat_{n_1, n_1} under the differential d is*

$$d(\text{Mat}_{n_1, n_1}) = \mathcal{L}[1]. \quad (3)$$

Proof. For any $A \in \text{Mat}_{n_1, n_1}$, let $T' = d(A, 0, 0, 0) = (A \otimes I \otimes I) \cdot T$. Consider the slices $L_i = [T'_{ijk}]_{j,k}$. Then, it is straightforward to compute that for $1 \leq i \leq n_1$, the slice

$$[T'_{ijk}]_{j,k} = \sum_{t=1}^{n_1} a_{it} L_i. \quad \blacktriangleleft$$

Similar to L_i , define the slices in the other directions as $M_j = [T'_{ijk}]_{i,k}$ and $N_k = [T'_{ijk}]_{i,j}$. Let \mathcal{M} and \mathcal{N} denote the spans of the M_j s and N_k s respectively.

► **Corollary 19.** *The image of the differential d is*

$$\text{im } d = \mathcal{L}[1] + \mathcal{M}[2] + \mathcal{N}[3] + Y_r.$$

Proof. This follows from Lemma 18 (and its counterparts in the other two directions) and Lemma 17. \blacktriangleleft

Now, we refine the above corollary. Denote by \widehat{L}_i the top-left $r \times r$ submatrix of L_i . Then, let $\widehat{\mathcal{L}} = \text{span}(\widehat{L}_1, \widehat{L}_2, \dots, \widehat{L}_{n_1})$. Similarly define $\widehat{M}_j, \widehat{N}_k$ and $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{N}}$.

► **Proposition 20.** *The image of the differential d is*

$$\text{im } d = \widehat{\mathcal{L}}[1] + \widehat{\mathcal{M}}[2] + \widehat{\mathcal{N}}[3] + Y_r.$$

Proof. This follows directly from Corollary 19 and the definition of the coordinate subspace Y_r . \blacktriangleleft

Now, we can finally prove Theorem 13.

Proof of Theorem 13. Let $T \in X_r$ be generic. Define L_i, M_i, N_i as above. Let

$$\mathcal{X}_1 = \text{span}(\widehat{L}_{r+1}, \widehat{L}_{r+2}, \dots, \widehat{L}_{n_1}).$$

Then \mathcal{X}_1 is a generic subspace of $\text{Mat}_{r,r}$ of dimension $n_1 - r$. We similarly define

$$\begin{aligned} \mathcal{X}_2 &= \text{span}(\widehat{M}_{r+1}, \widehat{M}_{r+2}, \dots, \widehat{M}_{n_2}) \\ \mathcal{X}_3 &= \text{span}(\widehat{N}_{r+1}, \widehat{N}_{r+2}, \dots, \widehat{N}_{n_3}). \end{aligned}$$

By hypothesis, $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{r,r,r}$. Hence, by Proposition 20, we see that

$$\text{im}(d) = \widehat{\mathcal{L}}[1] + \widehat{\mathcal{M}}[2] + \widehat{\mathcal{N}}[3] + Y_r \supseteq \mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + Y_r = K^{r,r,r} + Y_r = K^{n_1, n_2, n_3}.$$

Thus, we get that $Q(n_1, n_2, n_3) \geq r$ by Corollary 16. \blacktriangleleft

We observe that the idea from the proof of Theorem 13 actually yields the stronger version below. To state this we define the linear subspace

$$W_r = \{T \in K^{r,r,r} \mid T_{ijk} = 0 \text{ if } i, j, k \text{ are all different}\} \subseteq K^{r,r,r}.$$

► **Theorem 21.** *Let $r \leq n_1, n_2, n_3$ such that $n_i = r < r^2$ for $i = 1, 2, 3$. Let $\mathcal{X}_i \subseteq \text{Mat}_{r,r}$ be a generic subspace of dimension $n_i - r$ for $i = 1, 2, 3$. Suppose*

$$\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r},$$

then $Q(n_1, n_2, n_3) \geq r$. Further, if characteristic of K is 0, then we have the converse.

In other words, if characteristic of K is 0, then Theorem 21 says that $Q(n_1, n_2, n_3)$ is given by the smallest number r such that $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r}$ for generic subspaces $\mathcal{X}_i \subseteq \text{Mat}_{r,r}$ of dimension $n_i - r$.

Proof. The proof is similar to that of Theorem 13. Take a generic $T \in X_r$. Define L_i, M_i, N_i as above. Set $\mathcal{X}_1 = \text{span}(\widehat{L}_{r+1}, \widehat{L}_{r+2}, \dots, \widehat{L}_{n_1})$. Then \mathcal{X}_1 is a generic subspace of $\text{Mat}_{r,r}$ of dimension $n_1 - r$. Let $P_1 = \text{span}(\widehat{L}_1, \widehat{L}_2, \dots, \widehat{L}_r)$. Similarly define $\mathcal{X}_2 = \text{span}(\widehat{M}_{r+1}, \widehat{M}_{r+2}, \dots, \widehat{M}_{n_2})$ and $\mathcal{X}_3 = \text{span}(\widehat{N}_{r+1}, \widehat{N}_{r+2}, \dots, \widehat{N}_{n_3})$ and also $P_2 = \text{span}(\widehat{M}_1, \dots, \widehat{M}_r)$ and $P_3 = \text{span}(\widehat{N}_1, \dots, \widehat{N}_r)$.

It is a straightforward computation to see that since T is generic in X_r , we have $P_1[1] + P_2[2] + P_3[3] = W_r$. Hence, using Proposition 20, we see that

$$\begin{aligned} \text{im}(d) &= \widehat{\mathcal{L}}[1] + \widehat{\mathcal{M}}[2] + \widehat{\mathcal{N}}[3] + Y_r \\ &= \mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + P_1[1] + P_2[2] + P_3[3] + Y_r \\ &= \mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r + Y_r \end{aligned}$$

Now, we claim that

$$\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r} \Leftrightarrow \mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r + Y_r = K^{n_1, n_2, n_3}.$$

The proof of the claim is rather straightforward, the only subtle point being that Y_r and $K^{r,r,r}$ have an intersection (i.e., the main diagonal of $K^{r,r,r}$), but this intersection is included in W_r anyway, so the claim goes through.

Thus, we get that $\text{im}(d) = K^{n_1, n_2, n_3}$ if and only if $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r}$. Now, the theorem follows from applying Corollary 16. \blacktriangleleft

In the next section we will prove that the requirement of Theorem 13 is satisfied. This goes as follows. We will show (Lemma 26) that:

► **Lemma 22.** *Let $\mathcal{X}_i \subseteq \text{Mat}_{3,3}$ be a generic subspace of dimension 3 for $i = 1, 2, 3$. Then $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{3,3,3}$.*

Then from a blow-up argument we obtain:

► **Lemma 23.** *Let $\mathcal{X}_i \subseteq \text{Mat}_{3d,3d}$ be a generic subspace of dimension $3d^2$ for $i = 1, 2, 3$. Then $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{3d,3d,3d}$.*

Proof. It is enough to construct one such choice of \mathcal{X}_i . By Lemma 22 there exist subspaces $\mathcal{Y}_i \subseteq \text{Mat}_{3,3}$ of dimension 3 such that $\mathcal{Y}_1[1] + \mathcal{Y}_2[2] + \mathcal{Y}_3[3] = K^{3,3,3}$. Then $\mathcal{X}_i = \mathcal{Y}_i \otimes \text{Mat}_{d,d}$ satisfy the claim. \blacktriangleleft

► **Theorem 24.** *Let d be a positive integer such that $n - 3d \geq 3d^2$. Then $Q(n, n, n) \geq 3d$.*

Proof. Taking $r = 3d$, this follows from Theorem 13 and Lemma 23. \blacktriangleleft

Proof of Theorem 3. If we take $d = \lfloor \sqrt{n/3 + 1/4} - 1/2 \rfloor$, then we have $d \leq \sqrt{n/3 + 1/4} - 1/2$, so $(d + 1/2)^2 \leq n/3 + 1/4$. Thus, $d^2 + d + 1/4 \leq n/3 + 1/4$, so $d^2 + d \leq n/3$ or equivalently $3d^2 \leq n - 3d$. Thus, it follows from Theorem 24 that $Q(n) \geq 3d$. \blacktriangleleft

For the higher-order lower bound Theorem 6 we prove (Lemma 27) the analogous higher-order version of Lemma 22 which we can similarly blow up and apply Theorem 13 to.

4 Constructions of tensor space decompositions

We assume that the base field K is infinite.

Let \mathcal{C} be the set of all matrices $\begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ a_1 & a_2 & \cdots & a_d \end{bmatrix}$ for which n_1, n_2, \dots, n_d are positive integers, and a_1, a_2, \dots, a_d are nonnegative integers, and $\sum_{i=1}^d a_i n_i = \prod_{i=1}^d n_i$. Let \mathcal{S} be the subset of all matrices $\begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ a_1 & a_2 & \cdots & a_d \end{bmatrix} \in \mathcal{C}$ with the following property: If V_1, V_2, \dots, V_d are vector spaces of dimensions n_1, n_2, \dots, n_d respectively, and W_i is a general subspace of $\widehat{V}_i := V_1 \otimes V_2 \otimes \cdots \otimes V_{i-1} \otimes V_{i+1} \otimes \cdots \otimes V_d$ of dimension a_i for all i , then $\sum_{i=1}^d \Phi_i(W_i \otimes V_i) = V_1 \otimes \cdots \otimes V_d$, where $\Phi : \widehat{V}_i \otimes V_i \rightarrow V_1 \otimes V_2 \otimes \cdots \otimes V_d$ is the isomorphism given by permuting the factors.

4.1 General construction methods

We will use the following simple facts. It is obvious that if a matrix lies in \mathcal{S} and we permute its columns, then it will still lie in \mathcal{S} . It is also clear that $\begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ a_1 & a_2 & \cdots & a_d \end{bmatrix}$ lies in \mathcal{S} if and only if $\begin{bmatrix} n_1 & n_2 & \cdots & n_d & 1 \\ a_1 & a_2 & \cdots & a_d & 0 \end{bmatrix}$ lies in \mathcal{S} . The vector $\begin{bmatrix} n \\ 1 \end{bmatrix}$ lies in \mathcal{S} for all positive integers n . Finally,

$$\begin{bmatrix} n_1 & n_2 & \cdots & n_d & 1 & 1 \\ a_1 & a_2 & \cdots & a_d & b_1 & b_2 \end{bmatrix} \in \mathcal{S} \quad \text{if and only if} \quad \begin{bmatrix} n_1 & n_2 & \cdots & n_d & 1 \\ a_1 & a_2 & \cdots & a_d & b_1 + b_2 \end{bmatrix} \in \mathcal{S} \quad (4)$$

► **Lemma 25** (Direct sum construction). *Suppose that $n_d = n'_d + n''_d$, and $a_i = a'_i + a''_i$ for $i = 1, 2, \dots, d-1$, and*

$$\begin{bmatrix} n_1 & n_2 & \cdots & n_{d-1} & n'_d \\ a'_1 & a'_2 & \cdots & a'_{d-1} & a_d \end{bmatrix}, \begin{bmatrix} n_1 & n_2 & \cdots & n_{d-1} & n''_d \\ a''_1 & a''_2 & \cdots & a''_{d-1} & a_d \end{bmatrix} \in \mathcal{S}.$$

Then we have $\begin{bmatrix} n_1 & n_2 & \cdots & n_d \\ a_1 & a_2 & \cdots & a_d \end{bmatrix} \in \mathcal{S}$.

Proof. Suppose V_1, V_2, \dots, V_d are vector spaces of dimensions n_1, n_2, \dots, n_d respectively and choose a general subspace W_d of V_d of dimension a_d . We can write $V_d = V'_d \oplus V''_d$ where V'_d and V''_d have dimensions n'_d and n''_d respectively. For $i = 1, 2, \dots, d-1$ choose a general subspace $W'_i \subseteq V_1 \otimes \cdots \otimes V_{i-1} \otimes V_{i+1} \otimes \cdots \otimes V_{d-1} \otimes V'_d$ of dimension a'_i such that $V_1 \otimes V_2 \otimes \cdots \otimes V_{d-1} \otimes V'_d$ is equal to $(\sum_{i=1}^{d-1} \Phi_i(W'_i \otimes V_i)) + W_d \otimes V'_d$. Similarly, for $i = 1, 2, \dots, d-1$, choose a general subspace $W''_i \subseteq V_1 \otimes \cdots \otimes V_{i-1} \otimes V_{i+1} \otimes \cdots \otimes V_{d-1} \otimes V''_d$ of dimension a''_i such that $V_1 \otimes V_2 \otimes \cdots \otimes V_{d-1} \otimes V''_d$ is equal to $(\sum_{i=1}^{d-1} \Phi_i(W''_i \otimes V_i)) + W_d \otimes V''_d$. Set $W_i = W'_i \oplus W''_i \subseteq V_1 \otimes \cdots \otimes V_{i-1} \otimes V_{i+1} \otimes \cdots \otimes V_d$ for $i = 1, 2, \dots, d-1$. Then we have

$$\begin{aligned} \sum_{i=1}^d \Phi_i(W_i \otimes V_i) &= \left(\sum_{i=1}^{d-1} \Phi_i((W'_i \oplus W''_i) \otimes V_i) \right) + W_d \otimes (V'_d \oplus V''_d) = \\ &= \left(\left(\sum_{i=1}^{d-1} \Phi_i(W'_i \otimes V_i) \right) + W_d \otimes V'_d \right) \oplus \left(\left(\sum_{i=1}^{d-1} \Phi_i(W''_i \otimes V_i) \right) + W_d \otimes V''_d \right) = \\ &= (V_1 \otimes \cdots \otimes V_{d-1} \otimes V'_d) \oplus (V_1 \otimes \cdots \otimes V_{d-1} \otimes V''_d) = V_1 \otimes V_2 \otimes \cdots \otimes V_d. \end{aligned}$$

This finishes the proof. ◀

4.2 Recursive construction for every order

The following lemma gives Lemma 22 which we used in the proof of the generic subrank lower bound for order three. After that we will recursively obtain what is needed for the higher-order case generic subrank lower bound.

► **Lemma 26** (Base case). *We have $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix} \in \mathcal{S}$.*

Proof. We can verify explicitly that $\begin{bmatrix} 2 & 2 & 2 \\ 0 & 3 & 1 \end{bmatrix} \in \mathcal{S}$, as follows. For a 3-dimensional subspace W_1 of $V_2 \otimes V_3 = K^{2 \times 2}$, take the space of all matrices of the form

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

and for the 1-dimensional subspace W_3 of $V_1 \otimes V_2 = K^{2 \times 2}$, take the span of the identity matrix. A tensor in the intersection of $W_1 \otimes V_3 \cap V_1 \otimes W_3 \subseteq V_1 \otimes V_2 \otimes V_3$ is of the form

$$\left(\begin{array}{cc|cc} a_1 & b_1 & a_2 & b_2 \\ b_1 & c_1 & b_2 & c_2 \end{array} \right) = \left(\begin{array}{cc|cc} p_1 & 0 & 0 & p_1 \\ p_2 & 0 & 0 & p_2 \end{array} \right)$$

for some $a_1, a_2, b_1, b_2, c_1, c_2, p_1, p_2$. Clearly, $b_1 = c_1 = a_2 = b_2 = 0$. So we get $p_1 = b_2 = 0$ and $p_2 = b_1 = 0$. So $W_1 \otimes V_3 \cap V_1 \otimes W_3 = 0$ and $W_1 \otimes V_3 + V_1 \otimes W_3$ has dimension $3 \cdot 2 + 2 \cdot 1 = 8$ and therefore must be equal to $V_1 \otimes V_2 \otimes V_3 = K^{2 \times 2 \times 2}$. Using Lemma 25, we now deduce:

- $\begin{bmatrix} 2 & 2 & 2 \\ 0 & 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix} \in \mathcal{S}$, so $\begin{bmatrix} 3 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 2+1 & 2 & 2 \\ 0 & 3+0 & 1+2 \end{bmatrix} \in \mathcal{S}$
- $\begin{bmatrix} 3 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 0 \end{bmatrix} \in \mathcal{S}$, so $\begin{bmatrix} 3 & 1 & 2 \\ 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1+1 \\ 1+0 & 0+3 & 0 \end{bmatrix} \in \mathcal{S}$
- $\begin{bmatrix} 3 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 2 \\ 1 & 3 & 0 \end{bmatrix} \in \mathcal{S}$, so $\begin{bmatrix} 3 & 3 & 2 \\ 1 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 2+1 & 2 \\ 0+1 & 3 & 3+0 \end{bmatrix} \in \mathcal{S}$
- $\begin{bmatrix} 3 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix} \in \mathcal{S}$, so $\begin{bmatrix} 3 & 3 & 1 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 1+1+1 & 1 \\ 1+1+0 & 0 & 0+0+3 \end{bmatrix} \in \mathcal{S}$
- $\begin{bmatrix} 3 & 3 & 1 \\ 2 & 0 & 3 \end{bmatrix}, \begin{bmatrix} 3 & 3 & 2 \\ 1 & 3 & 3 \end{bmatrix} \in \mathcal{S}$, so $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 1+2 \\ 2+1 & 0+3 & 3 \end{bmatrix} \in \mathcal{S}$

The final line finishes the proof. ◀

To continue, we define the “concatenation” notation:

$$\begin{bmatrix} n_1 & n_2 & \dots & n_d \\ a_1 & a_2 & \dots & a_d \end{bmatrix} \odot \begin{bmatrix} m_1 & m_2 & \dots & m_e \\ b_1 & b_2 & \dots & b_e \end{bmatrix} = \begin{bmatrix} n_1 & n_2 & \dots & n_d & m_1 & m_2 & \dots & m_e \\ a_1 & a_2 & \dots & a_d & b_1 & b_2 & \dots & b_e \end{bmatrix}$$

and the k -fold repeated version of this notation:

$$\begin{bmatrix} n_1 & n_2 & \dots & n_d \\ a_1 & a_2 & \dots & a_d \end{bmatrix}^{\odot k} = \underbrace{\begin{bmatrix} n_1 & n_2 & \dots & n_d \\ a_1 & a_2 & \dots & a_d \end{bmatrix} \odot \begin{bmatrix} n_1 & n_2 & \dots & n_d \\ a_1 & a_2 & \dots & a_d \end{bmatrix} \odot \dots \odot \begin{bmatrix} n_1 & n_2 & \dots & n_d \\ a_1 & a_2 & \dots & a_d \end{bmatrix}}_k$$

We will now use Lemma 25 and Lemma 26 to prove:

► **Lemma 27.** *For all $n \geq 3$, we have $\begin{bmatrix} n \\ n^{n-2} \end{bmatrix}^{\odot n} \in \mathcal{S}$.*

Proof. We already know the case $n = 3$ (Lemma 26), so assume that $n \geq 4$.

We will first show that $\begin{bmatrix} n \\ n \end{bmatrix}^{\odot 3} \odot \begin{bmatrix} 1 \\ n^2 \end{bmatrix}^{\odot (n-3)} \in \mathcal{S}$. We note that a straightforward direct construction gives

$$\begin{bmatrix} n & n & 1 \\ k & 0 & n^2 - kn \end{bmatrix} \in \mathcal{S}$$

for $k = 1, 2, 3$. To proceed, we choose nonnegative integers $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n \in \{0, 1, 2, 3\}$ such that $a_1 + a_2 + \dots + a_n = n$, $b_1 + b_2 + \dots + b_n = n$ and $a_i b_i = 0$ for all i . Indeed this can be done as follows: If $n = 2m$ is even, then we can take $a_1 = a_2 = \dots = a_m = b_{m+1} = b_{m+2} = \dots = b_n = 2$ and $b_1 = b_2 = \dots = b_m = a_{m+1} = a_{m+2} = \dots = a_n = 0$. If $n = 2m + 1$ is odd, then we take $a_1 = 3$, $b_{m+1} = 1$, $a_2 = a_3 = \dots = a_m = b_{m+2} = b_{m+3} = \dots = b_n = 2$ and $b_1 = b_2 = \dots = b_m = a_{m+1} = a_{m+2} = \dots = a_n = 0$. Because $a_i b_i = 0$ we get that

$$\begin{bmatrix} n & n & 1 \\ a_i & b_i & n^2 - (a_i + b_i)n \end{bmatrix} \in \mathcal{S}$$

and then using (4) we get

$$\left[\begin{matrix} n & n & 1 \\ a_i & b_i & n \end{matrix} \begin{matrix} 1 \\ n^2 - (a_i + b_i + 1)n \end{matrix} \right] \in \mathcal{S} \tag{5}$$

for $i = 1, 2, \dots, n$. Applying the direct sum construction Lemma 25 to the elements of (5) we get

$$\left[\begin{matrix} n & n & n \\ n & n & n \end{matrix} \begin{matrix} 1 \\ n^3 - 3n^2 \end{matrix} \right] = \left[\sum_i^n a_i \sum_i^n b_i \begin{matrix} 1 + \dots + 1 \\ n \end{matrix} \sum_i n^2 - (a_i + b_i + 1)n \right] \in \mathcal{S}.$$

Applying (4) repeatedly to this we obtain $\left[\begin{matrix} n \\ n \end{matrix} \right]^{\odot 3} \odot \left[\begin{matrix} 1 \\ n^2 \end{matrix} \right]^{\odot (n-3)} \in \mathcal{S}$.

We will now show by induction on d that

$$\left[\begin{matrix} n \\ n^{d-2} \end{matrix} \right]^{\odot d} \odot \left[\begin{matrix} 1 \\ n^{d-1} \end{matrix} \right]^{\odot (n-d)} \in \mathcal{S} \tag{6}$$

for $d = 3, 4, \dots, n$. We already know that the base case $d = 3$ is true. Suppose for the induction step that $\left[\begin{matrix} n \\ n^{d-2} \end{matrix} \right]^{\odot d} \odot \left[\begin{matrix} 1 \\ n^{d-1} \end{matrix} \right]^{\odot (n-d)} \in \mathcal{S}$. Then we have using (4) that

$$\left[\begin{matrix} n \\ n^{d-1} \end{matrix} \right]^{\odot (d+1)} \odot \left[\begin{matrix} 1 \\ n^d \end{matrix} \right]^{\odot (n-d-1)} = \left[\begin{matrix} n & n & n \\ n^{d-2} & n & n \end{matrix} \right]^{\odot d} \odot \left[\begin{matrix} 1 + \dots + 1 \\ n^{d-1} \end{matrix} \right] \odot \left[\begin{matrix} 1 \\ n^{d-1} \end{matrix} \right]^{\odot (n-d)-1} \in \mathcal{S}.$$

This proves (6).

Finally, by setting $d = n$ in (6) we obtain the claim $\left[\begin{matrix} n \\ n^{n-2} \end{matrix} \right]^{\odot n} \in \mathcal{S}$. ◀

5 Application: Subrank is not additive

We discuss in this section an application of our upper bound Theorem 2 on the generic subrank, namely that the subrank is not additive under the direct sum. That is, there are tensors S, T such that $Q(T) + Q(S) < Q(T \oplus S)$. In fact, we obtain a large gap between $Q(T) + Q(S)$ and $Q(T \oplus S)$. The proof relies on the idea of writing the diagonal tensor I_n as a sum of two generic tensors and on basic properties of the subrank.¹¹

► **Theorem 28.** *There are tensors $S, T \in K^{n,n,n}$ such that we have $Q(T), Q(S) \leq \sqrt{3n-2}$ and $Q(T \oplus S) \geq n$.*

Proof. By Theorem 2 there is a non-empty Zariski-open subset $U \subseteq K^{n,n,n}$ such that for all $T \in U$ we have $Q(T) \leq \sqrt{3n-2}$. Recall that $I_n \in K^{n,n,n}$ is the tensor with ones in the diagonal entries and zeroes elsewhere. The subset $I_n - U \subseteq K^{n,n,n}$ is also non-empty and Zariski-open, and thus the intersection $U \cap (I_n - U)$ is non-empty. This means that there is a $T \in U$ such that $I_n - T$ is in U . Fix this T and let $S = I_n - T$. Then $Q(T), Q(S) \leq \sqrt{3n-2}$. For their direct sum, we observe the simple general fact that $T \oplus S \geq T + S$ where the left-hand side is the direct sum and the right-hand side is the coordinate-wise sum. Since subrank is monotone under \geq we find that $Q(T \oplus S) \geq Q(T + S) = Q(I_n) = n$. ◀

The proof of Theorem 28 extends directly so that similarly from Theorem 5 we get the following higher-order non-additivity result.

► **Theorem 29.** *There are k -tensors $S, T \in K^{n,\dots,n}$ such that $Q(T), Q(S) \leq (kn - (k-1))^{\frac{1}{k-1}}$ while $Q(T \oplus S) \geq n$.*

¹¹This idea was used earlier in [9] to show that the “lower support functionals” [41] are not additive. Their results can be used to find a weaker non-additivity for subrank (with a smaller gap, and only for tensors of order three), namely that there are tensors $S, T \in K^{n,n,n}$ such that $Q(T), Q(S) \leq n^{2/3+o(1)}$ and $Q(T \oplus S) \geq n$.

6 Open problems

There are several natural open problems that arise from or are closely related to our study and results on the generic subrank in this paper. We briefly list some of these problems here.

- While we determine the generic subrank very precisely up to a small additive constant, it is natural to ask whether our upper bound on the generic subrank $Q(n)$ is exactly tight. We know that this is the case for all small cases ($n \leq 100$).
- Closely related to the above, but for higher-order and unbalanced formats, what is the exact value of the generic subrank $Q(n_1, \dots, n_k)$ in K^{n_1, \dots, n_k} when the dimensions n_i are not all equal? In particular, is our upper bound tight?
- To attack the previous two problems the natural approach would be to use the stronger Theorem 21. Our current lower bound on the generic subrank uses Theorem 13 which relies on our constructions of decompositions $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] = K^{r,r,r}$. Theorem 21 suggests to instead construct decompositions $\mathcal{X}_1[1] + \mathcal{X}_2[2] + \mathcal{X}_3[3] + W_r = K^{r,r,r}$. What are the “best” constructions that can be obtained of this form?
- There is a natural approximative version of subrank called *border subrank*¹² which plays a central role in algebraic complexity theory (in particular the study of matrix multiplication algorithms). The border subrank is at least the subrank. What value does the border subrank take on generic tensors?
- What is the value of the generic asymptotic subrank $\underline{Q}(n)$? The state of the art is that $n^{2/3} \leq \underline{Q}(n) \leq n$, so in particular it remains open whether generic tensors have “full” generic asymptotic subrank or not. We note that we do know of explicit tensors for which the asymptotic subrank is not full (the so-called W-tensor, for example) and also of explicit tensors for which the asymptotic subrank is full in a non-trivial way (matrix multiplication tensors, for example). The aforementioned lower bound of $n^{2/3}$ is by Strassen’s elegant construction (which makes use of the matrix multiplication tensors). Our results show that one cannot obtain better lower bound on $\underline{Q}(n)$ by improving the lower bound on the generic subrank $Q(n)$ since $Q(n) = \theta(\sqrt{n})$.
- With the notation we introduced in Section 4 on constructions of tensor space decompositions, it is natural to ask what precisely are the elements of \mathcal{S} . Many elements can be constructed from simple base cases and the direct sum construction. However, we do not know whether a finite number of generators in this sense suffices to generate all of \mathcal{S} .
- We found tensors S, T for which there is a large gap between $Q(S \oplus T)$ and $Q(S) + Q(T)$. Is this the largest possible gap? More speculatively, we may ask: is there a general relation between direct sum problems (i.e. additivity under direct sum) and parameter values at generic instances?

References

- 1 Josh Alman. Limits on the universal method for matrix multiplication. In *34th Computational Complexity Conference (CCC 2019)*, pages 12:1–12:24, 2019. doi:10.4230/LIPIcs.CCC.2019.12.
- 2 Abhishek Bhowmick and Shachar Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory, 2015. arXiv:1506.02047.

¹²The difference with subrank is that the restriction preorder is replaced by the *degeneration* preorder, which results in the border subrank $\underline{Q}(T)$ of T being the largest number r such that L_r is in the $\mathrm{GL}_n^{\times 3}$ -orbit closure of T .

- 3 Abhishek Bhruhundi, Prahladh Harsha, Pooya Hatami, Swastik Kopparty, and Mrinal Kumar. On multilinear forms: Bias, correlation, and tensor rank. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*, pages 29:1–29:23, 2020. doi:10.4230/LIPIcs.APPROX/RANDOM.2020.29.
- 4 Markus Bläser. *Fast Matrix Multiplication*. Number 5 in Graduate Surveys. Theory of Computing Library, 2013. doi:10.4086/toc.gs.2013.005.
- 5 Markus Bläser. Explicit tensors. In *Perspectives in computational complexity*, pages 117–130. Springer, 2014.
- 6 Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, Eric Naslund, William F. Sawin, and Chris Umans. On cap sets and the group-theoretic approach to matrix multiplication. *Discrete Anal.*, pages Paper No. 3, 27, 2017. doi:10.19086/da.1245.
- 7 Armand Borel. *Linear algebraic groups*, volume 126 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991. doi:10.1007/978-1-4612-0941-6.
- 8 Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic complexity theory*, volume 315 of *Grundlehren Math. Wiss.* Springer-Verlag, Berlin, 1997. doi:10.1007/978-3-662-03338-8.
- 9 Peter Bürgisser. *Degenerationsordnung und Trägerfunktional bilinearer Abbildungen*. PhD thesis, Universität Konstanz, 1990. URL: <http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-20311>.
- 10 Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Universal points in the asymptotic spectrum of tensors (extended abstract). In *Proceedings of the 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC 2018)*, 2018. doi:10.1145/3188745.3188766.
- 11 Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Barriers for fast matrix multiplication from irreversibility. In *34th Computational Complexity Conference (CCC 2019)*, pages 26:1–26:17, 2019. doi:10.4230/LIPIcs.CCC.2019.26.
- 12 Alex Cohen and Guy Moshkovitz. Structure vs. randomness for bilinear maps. In *53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2021)*, 2021. doi:10.1145/3406325.3451007.
- 13 Alex Cohen and Guy Moshkovitz. Partition rank and analytic rank are uniformly equivalent, 2021. arXiv:2102.10509.
- 14 Harm Derksen. The G-stable rank for tensors, 2020. arXiv:2002.08435.
- 15 Klim Efremenko, Ankit Garg, Rafael Mendes de Oliveira, and Avi Wigderson. Barriers for rank methods in arithmetic complexity. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, pages 1:1–1:19, 2018. doi:10.4230/LIPIcs.ITCS.2018.1.
- 16 Jordan S. Ellenberg and Dion Gijswijt. On large subsets of \mathbb{F}_q^n with no three-term arithmetic progression. *Ann. of Math. (2)*, 185(1):339–343, 2017. doi:10.4007/annals.2017.185.1.8.
- 17 Ankit Garg, Visu Makam, Rafael Mendes de Oliveira, and Avi Wigderson. More barriers for rank methods, via a “numeric to symbolic” transfer. In *60th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2019)*, pages 824–844, 2019. doi:10.1109/FOCS.2019.00054.
- 18 Runshi Geng and J. M. Landsberg. On the geometry of geometric rank, 2021. arXiv:2012.04679.
- 19 W. T. Gowers. The slice rank of a direct sum, 2021. arXiv:2105.08394.
- 20 W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on \mathbb{F}_p^n . *Geom. Funct. Anal.*, 21(1):36–69, 2011. doi:10.1007/s00039-010-0106-3.
- 21 Ben Joseph Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contributions Discret. Math.*, 4(2), 2009. URL: <http://cdm.ucalgary.ca/cdm/index.php/cdm/article/view/133>.
- 22 Oliver Janzer. Low analytic rank implies low partition rank for tensors, 2018. arXiv:1809.10931.
- 23 Oliver Janzer. Polynomial bound for the partition rank vs the analytic rank of tensors. *Discrete Anal.*, 2020. arXiv:1902.11207.

- 24 Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pages 166–175, 2008. doi:10.1109/FOCS.2008.17.
- 25 Pascal Koiran. On tensor rank and commuting matrices, 2020. arXiv:2006.02374.
- 26 Swastik Kopparty, Guy Moshkovitz, and Jeroen Zuiddam. Geometric rank of tensors and subrank of matrix multiplication. In *35th Computational Complexity Conference (CCC 2020)*, 2020. doi:10.4230/LIPIcs.CCC.2020.35.
- 27 J. M. Landsberg and Giorgio Ottaviani. Equations for secant varieties of Veronese and other varieties. *Ann. Mat. Pura Appl. (4)*, 192(4):569–606, 2013. doi:10.1007/s10231-011-0238-6.
- 28 Shachar Lovett. The analytic rank of tensors and its applications. *Discrete Anal.*, 2019. doi:10.19086/da.8654.
- 29 Luka Milićević. Polynomial bound for partition rank in terms of analytic rank. *Geom. Funct. Anal.*, 29(5):1503–1530, 2019. doi:10.1007/s00039-019-00505-4.
- 30 Eric Naslund. The partition rank of a tensor and k -right corners in \mathbb{F}_q^n . *Journal of Combinatorial Theory, Series A*, 174:105190, 2020. doi:10.1016/j.jcta.2019.105190.
- 31 Eric Naslund and Will Sawin. Upper bounds for sunflower-free sets. *Forum Math. Sigma*, 5:Paper No. e15, 10, 2017. doi:10.1017/fms.2017.12.
- 32 Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *Journal of the ACM (JACM)*, 60(6):1–15, 2013.
- 33 Alexander A. Razborov. On submodular complexity measures. In *Proceedings of the London Mathematical Society symposium on Boolean function complexity*, pages 76–83, 1992. URL: <https://dl.acm.org/doi/10.5555/167687.167709>.
- 34 Robert Robere and Jeroen Zuiddam. Amortized circuit complexity, formal complexity measures, and catalytic algorithms. *Electron. Colloquium Comput. Complex.*, page 35, 2021. URL: <https://eccc.weizmann.ac.il/report/2021/035>.
- 35 Arnold Schönhage. Partial and total matrix multiplication. *SIAM J. Comput.*, 10(3):434–455, 1981. doi:10.1137/0210032.
- 36 Yaroslav Shitov. Counterexamples to Strassen’s direct sum conjecture. *Acta Math.*, 222(2):363–379, 2019. doi:10.4310/ACTA.2019.v222.n2.a3.
- 37 Volker Strassen. Vermeidung von Divisionen. *J. Reine Angew. Math.*, 264:184–202, 1973.
- 38 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.
- 39 Volker Strassen. Relative bilinear complexity and matrix multiplication. *J. Reine Angew. Math.*, 375/376:406–443, 1987. doi:10.1515/crll.1987.375-376.406.
- 40 Volker Strassen. The asymptotic spectrum of tensors. *J. Reine Angew. Math.*, 384:102–152, 1988. doi:10.1515/crll.1988.384.102.
- 41 Volker Strassen. Degeneration and complexity of bilinear maps: some asymptotic spectra. *J. Reine Angew. Math.*, 413:127–180, 1991. doi:10.1515/crll.1991.413.127.
- 42 Terence Tao. A symmetric formulation of the Croot-Lev-Pach-Ellenberg-Gijswijt capset bound. URL: <https://terrytao.wordpress.com>, 2016.
- 43 Avi Wigderson and Jeroen Zuiddam. Asymptotic spectra: Theory, applications and extensions, 2021. URL: <https://staff.fnwi.uva.nl/j.zuiddam/papers/convexity.pdf>.
- 44 Jeroen Zuiddam. *Asymptotic spectra, algebraic complexity and moment polytopes*. PhD thesis, University of Amsterdam, 2018.