THEORY OF COMPUTING, Volume 17 (1), 2021, pp. 1–30
www.theoryofcomputing.org

SPECIAL ISSUE: CCC 2019

Limits on the Universal Method for Matrix Multiplication

Josh Alman*

Received July 24, 2019; Revised July 19, 2020; Published September 3, 2021

Abstract. We prove limitations on the known methods for designing matrix multiplication algorithms. Alman and Vassilevska Williams (FOCS'18) recently defined the *Universal Method*, which generalizes all the known approaches, including Strassen's Laser Method (J. reine angew. Math., 1987) and Cohn and Umans's Group Theoretic Method (FOCS'03). We prove concrete lower bounds on the algorithms one can design by applying the Universal Method to many different tensors. Our proofs use new tools to give upper bounds on the *asymptotic slice rank* of a wide range of tensors. Our main result is that the Universal Method applied to any Coppersmith–Winograd tensor CW_q cannot yield a bound on ω, the exponent of matrix multiplication, better than 2.16805. It was previously known (Alman and Vassilevska Williams, FOCS'18) that the weaker "Galactic Method" applied to CW_q could not achieve an exponent of 2.

We also study the Laser Method (which is a special case of the Universal Method) and prove that it is "complete" for matrix multiplication algorithms: when it applies to a tensor T, it achieves $\omega = 2$ if and only if it is possible for the Universal Method applied to T to achieve $\omega = 2$. Hence, the Laser Method, which was originally used as an algorithmic tool,

A conference version of this paper appeared in the 34th Computational Complexity Conference, 2019 [1].

*Supported in part by NSF grants CCF-1651838 and CCF-1741615. The work reported here was done while the author was a Ph. D. student at MIT.

ACM Classification: F.2.2

AMS Classification: 68Q17

Key words and phrases: matrix multiplication, algebraic complexity

DOI: 10.4086/toc.2021.v017a001

can also be seen as a lower-bound tool for a large class of algorithms. For example, in their landmark paper, Coppersmith and Winograd (J. Symbolic Computation, 1990) achieved a bound of $\omega \leq 2.376$, by applying the Laser Method to CW_q. By our result, the fact that they did not achieve $\omega = 2$ *implies* a lower bound on the Universal Method applied to CW_q.

1 Introduction

One of the biggest open questions in computer science asks how quickly one can multiply two matrices. Progress on this problems is measured by giving bounds on ω, the *exponent of matrix multiplication*, defined as the smallest real number such that two $n \times n$ matrices over a field can be multiplied using $n^{\omega+o(1)}$ field operations. Since Strassen's breakthrough algorithm [30] showing that $\omega \leq \log_2(7) \approx 2.81$, there has been a long line of work, resulting in the current best bound of $\omega \leq 2.37286$ [36, 26, 4], and it is popularly conjectured that $\omega = 2$ (see, e. g., [19, 15, 17, 5] which present conjectures that would imply $\omega = 2$).

The key to Strassen's algorithm is an algebraic identity showing how $2 \times 2 \times 2$ matrix multiplication can be computed surprisingly efficiently (in particular, Strassen showed that the $2 \times 2 \times 2$ matrix multiplication tensor has rank at most 7; see Section 2 for precise definitions). Arguing about the ranks of larger matrix multiplication tensors has proven to be quite difficult—in fact, even the rank of the $3 \times 3 \times 3$ matrix multiplication tensor isn't currently known. Progress on bounding ω since Strassen's algorithm has thus taken the following approach: Pick a tensor (trilinear form) T, typically not a matrix multiplication tensor, such that

- Powers $T^{\otimes n}$ of T can be "efficiently computed" (i. e., T has low asymptotic rank), and
- T is useful for performing matrix multiplication, since large matrix multiplication tensors can be "embedded" within powers of T.

Combined, these give an upper bound on the rank of matrix multiplication itself, and hence ω.

The most general type of embedding which is known to preserve the ranks of tensors as required for the above approach is a *degeneration*. In [3], Vassilevska Williams and the author called this method of taking a tensor T and finding the best possible degeneration of powers $T^{\otimes n}$ into matrix multiplication tensors the *Universal Method applied to* T, and the best bound on ω which can be proved in this way is written $\omega_u(T)$. They also defined two weaker methods: *the Galactic Method applied to* T, in which the "embedding" must be a more restrictive *monomial degeneration*, resulting in the bound $\omega_g(T)$ on ω, and *the Solar Method applied to* T, in which the "embedding" must be an even more restrictive *zeroing out*, resulting in the bound $\omega_s(T)$ on ω. Since monomial degenerations and zeroing outs are successively more restrictive types of degenerations, we have that for all tensors T,

$$\omega \leq \omega_u(T) \leq \omega_g(T) \leq \omega_s(T).$$

These methods are *very general*; there are no known methods for computing $\omega_u(T)$, $\omega_g(T)$, or $\omega_s(T)$ for a given tensor T, and these quantities are even unknown for very well-studied tensors T. The two main approaches to designing matrix multiplication algorithms are the Laser Method of Strassen [32] and the Group-Theoretic Method of Cohn and Umans [16]. Both of these approaches show how to give upper bounds on $\omega_s(T)$ for particular structured tensors T (and hence upper bound ω itself). In other words, they both give ways to find zeroing outs of tensors into matrix multiplication tensors, but not necessarily the best zeroing outs. In fact, it is known that the Laser Method does not always give the best zeroing out for a particular tensor T, since the improvements from [19] to later works [21, 36, 26, 4] can be seen as giving slight improvements to the Laser Method to find better and better zeroing outs.[1] The Group-Theoretic Method, like the Solar Method, is very general, and it is not clear how to optimally apply it to a particular group or family of groups.

All of the improvements on bounding ω for the past 30+ years have come from studying the Coppersmith–Winograd family of tensors $\{CW_q\}_{q\in\mathbb{N}}$. The Laser Method applied to powers of CW_5 gives the bound $\omega_s(CW_5) \leq 2.3729$. The Group-Theoretic Method can also prove the best known bound $\omega \leq 2.3729$, by simulating the Laser Method analysis of CW_q (see, e. g., [2] for more details). Despite a long line of work on matrix multiplication, there are no known tensors[2] which seem to come close to achieving the bounds one can obtain using CW_q. This leads to the first main question of this paper:

Question 1.1. How much can we improve our bound on ω using a more clever analysis of the Coppersmith–Winograd tensor?

Vassilevska Williams and the author [3] addressed this question by showing that there is a constant $c > 2$ so that for all q, $\omega_g(CW_q) > c$. In other words, the Galactic Method (monomial degenerations) cannot be used with CW_q to prove $\omega = 2$. However, this leaves open a number of important questions: How close to 2 can we get using monomial degenerations; could it be that $\omega_g(CW_q) \leq 2.1$? Perhaps more importantly, what if we are allowed to use arbitrary degenerations; could it be that $\omega_u(CW_q) \leq 2.1$, or even $\omega_u(CW_q) = 2$?

The second main question of this paper concerns the Laser Method. The Laser Method computes an upper bound on $\omega_s(T)$ for any tensor T with certain structure (which we describe in detail in Section 5), and has led to every improvement on ω since its introduction by Strassen [32].

Question 1.2. When the Laser Method applies to a tensor T, how close does it come to optimally analyzing T?

As discussed, we know the Laser Method does not always give a tight bound on $\omega_s(T)$. For instance, Coppersmith–Winograd [19] applied the Laser Method to CW_q to prove $\omega_s(CW_q) \leq 2.376$, and then later work [21, 36, 26, 4] analyzed higher and higher powers of CW_q to show $\omega_s(CW_q) \leq 2.373$. Ambainis, Filmus and Le Gall [6] showed that analyzing higher and higher powers of CW_q itself with the Laser Method cannot yield an upper bound better than $\omega_s(CW_q) \leq 2.3725$. What about for other tensors? Could there be a tensor such that applying the Laser Method to T yields $\omega_s(T) \leq c$ for some $c > 2$, but applying the Laser Method to high powers $T^{\otimes n}$ of T yields $\omega_s(T) = 2$? Could applying an entirely different method to such a T, using arbitrary degenerations and not just zeroing outs, show that $\omega_u(T) = 2$?

1.1 Our results

We give strong resolutions to both Question 1.1 and Question 1.2.

[1]These works apply the Laser Method to higher powers of the tensor $T = CW_q$, a technique which is still captured by the Solar Method.

[2]Vassilevska Williams and the author [3] study a generalization of CW_q which can tie the best known bound, but its analysis is identical to that of CW_q. Our lower bounds in this paper will apply equally well to this generalized class as to CW_q itself.

Universal Method lower bounds. To resolve Question 1.1, we prove a new lower bound for the Coppersmith–Winograd tensor:

Theorem 1.3. *$\omega_u(CW_q) \geq 2.16805$ for all q.*

In other words, no analysis of CW_q, using any techniques within the Universal Method, can prove a bound on ω better than 2.16805. This generalizes the main result of [3] from the Galactic Method to the *Universal* Method, and gives a more concrete lower bound, increasing the bound from "a constant greater than 2" to 2.16805. We also give stronger lower bounds for particular tensors in the family. For instance, for the specific tensor CW_5 which yields the current best bound on ω, we show $\omega_u(CW_5) \geq 2.21912...$

Our techniques for proving lower bounds on $\omega_u(T)$ apply to CW_q as well as many other tensors of interest. We thus prove many other lower bounds, including: the same lower bound $\omega_u(CW_{q,\sigma}) \geq 2.16805$ for any *generalized Coppersmith–Winograd tensor* $CW_{q,\sigma}$ as introduced in [3], a similar lower bound for $cw_{q,\sigma}$, the generalized "simple" Coppersmith–Winograd tensor missing its "corner terms," and a lower bound for T_q, the structural tensor of the cyclic group C_q, matching the lower bounds obtained by [2, 9]. In Section 4 we give tables of our precise lower bounds for these and other tensors. Furthermore, we will show that our new tools for proving lower bounds on $\omega_u(T)$ are able to prove at least as high a lower bound as the tools of [3] can prove on $\omega_g(T)$. In other words, all of those previously known Galactic Method lower bounds hold for the Universal Method as well.

We also show how our lower-bound techniques can be used to study other properties of tensors. Coppersmith and Winograd [19] introduced the notion of the *value* $V_\tau(T)$ of a tensor T, which is useful when applying the Laser Method to a larger tensor T' which contains T as a subtensor. We show how our slice rank lower-bound tools yield a tight upper bound on the value of t_{112}, the notorious subtensor of $CW_q^{\otimes 2}$ which arises when applying the Laser Method to powers of CW_q. Although the value $V_\tau(t_{112})$ appears in every analysis of CW_q since [19], including [21, 36, 26, 4, 25, 27], the best lower bound on it has not improved since [19], and our new upper bound here helps explain why. See Subsection 2.5 and Subsection 4.4 for more details.

We briefly note that our lower bound of $2.16805 > 2 + \frac{1}{6}$ in Theorem 1.3 may be significant when compared to the recent algorithm of Cohen, Lee and Song [14] which solves n-variable linear programs in time about $O(n^\omega + n^{2+1/6})$.

The Laser Method is "complete." We next study in more detail how our results apply to tensors which we call *laser-ready* tensors. These are the tensors to which the Laser Method (as used by [19] on CW_q) applies; see Definition 5.1 for the precise definition. Tensors need certain structure to be laser-ready, but tensors T with this structure are essentially the only ones for which upper-bound techniques for $\omega_u(T)$ are known. In fact, every record-holding tensor in the history of matrix multiplication algorithm design has been laser-ready.

We will see below that many of the same properties of a tensor T which we use in this paper to prove lower bounds on $\omega_u(T)$ are also used in the Laser Method to give upper bounds on $\omega_s(T)$ when T is laser-ready. In particular, this connection will give an intriguing answer to Question 1.2:

Theorem 1.4. *If T is a laser-ready tensor, and $\omega_u(T) = 2$, then the Laser Method applied to T yields the bound $\omega_u(T) = 2$.*

In other words: If T is any tensor to which the Laser Method applies (as in Definition 5.1), and the Laser Method does not yield $\omega = 2$ when applied to T, then in fact $\omega_u(T) > 2$, and even the substantially more general Universal Method applied to T cannot yield $\omega = 2$. Hence, the Laser Method, which was originally used as an algorithmic tool, can also be seen as a lower-bound tool. Conversely, Theorem 1.4 shows that the Laser Method is "complete," in the sense that it cannot yield a bound on ω worse than 2 when applied to a tensor which is able to prove $\omega = 2$.

Theorem 1.4 explains and generalizes a number of phenomena:

- The fact that Coppersmith–Winograd [19] applied the Laser Method to the tensor CW_q and achieved an upper bound greater than 2 on ω *implies that* $\omega_u(CW_q) > 2$, and no *arbitrary degeneration* of powers of CW_q can yield $\omega = 2$.

- As mentioned above, it is known that applying the Laser Method to higher and higher powers of a tensor T can successively improve the resulting upper bound on ω. Theorem 1.4 shows that if the Laser Method applied to the first power of any tensor T did not yield $\omega = 2$, then this sequence of Laser Method applications (which is a special case of the Universal Method) must converge to a value greater than 2 as well. This generalizes the result of Ambainis, Filmus and Le Gall [6], who proved this about applying the Laser Method to higher and higher powers of the specific tensor $T = CW_q$.

- Our result also generalizes the result of Kleinberg, Speyer and Sawin [24] which studied the tensor T_q^{lower}, the lower triangular part of T_q. Kleinberg, Speyer and Sawin studied the *asymptotic slice rank* of T_q^{lower}, a quantity which will be featured prominently in our proofs below. They showed that (what can be seen as) the Laser Method achieves a tight lower bound on the asymptotic slice rank, matching an upper bound of Blasiak et al. [9]. Indeed, since T_q^{lower} is a laser-ready tensor, our proof will also imply this.

1.2 Overview of our techniques

We give a brief overview of the techniques we use to prove our main results, Theorem 1.3 and Theorem 1.4. All the technical terms we refer to here will be precisely defined in Section 2.

Section 2.6: Asymptotic slice rank and its connection with matrix multiplication. The tensors we study are 3-tensors, which can be seen as trilinear forms over three sets X,Y,Z of formal variables. A key notion in all our proofs will be the slice rank $\mathrm{S}(T)$ of a tensor T. $\mathrm{S}(T)$ is a measure of the complexity of T, analogous to the rank of a matrix. The notion of slice rank was first introduced by Tao [34], and then used by Blasiak et al. [9] in the context of lower bounds against the Group-Theoretic Method. In order to study degenerations of *powers* of tensors, rather than just tensors themselves, we study in this paper the *asymptotic slice rank* $\tilde{\mathrm{S}}(T)$ of tensors T:

$$\tilde{\mathrm{S}}(T) := \sup_{n\in\mathbb{N}} \mathrm{S}(T^{\otimes n})^{1/n}.$$

$\tilde{\mathrm{S}}$ satisfies two key properties:

1. Degenerations cannot increase the asymptotic slice rank of a tensor. In other words, if A degenerates to B, then $\tilde{S}(B) \le \tilde{S}(A)$.

2. Matrix multiplication tensors have high asymptotic slice rank.

This means that if a certain tensor T has a small value of $\omega_u(T)$, or in other words, powers $T^{\otimes n}$ can degenerate into large matrix multiplication tensors, then T itself must have large asymptotic slice rank. Hence, in order to lower bound $\omega_u(T)$, it suffices to upper bound $\tilde{S}(T)$.

Section 3: Tools for proving upper bounds on asymptotic slice rank. In general, bounding $\tilde{S}(T)$ for a tensor T can be much more difficult than bounding $S(T)$. This is because S can be supermultiplicative, i. e., there are tensors A and B such that $S(A) \cdot S(B) \ll S(A \otimes B)$. Indeed, we will show that $\tilde{S}(T) > S(T)$ for many tensors T of interest, including $T = CW_q$.

We will give three new upper-bound tools for $\tilde{S}(T)$ for many tensors T. Each applies to tensors with different properties:

- Theorem 3.2: If T is over X,Y,Z, then it is straightforward to see that if one of the sets of variables is not too large, then $\tilde{S}(T)$ must be small: $\tilde{S}(T) \le \min\{|X|,|Y|,|Z|\}$. In this first tool, we show how if T can be written as a sum $T = T_1 + \cdots + T_k$ of a few tensors, and each T_i does not have many of one type of variable, then we can still derive an upper bound on $\tilde{S}(T)$.

- Theorem 3.4: The second tool concerns partitions of the sets of variables X,Y,Z. It shows that if $\tilde{S}(T)$ is large, then there is a probability distribution on the blocks of T (subtensors corresponding to a choice of one part from each of the three partitions) so that the total probability mass assigned to each part of each partition is proportional to its size. Loosely, this means that T must have many different "symmetries," no matter how its variables are partitioned.

- Theorem 3.9: Typically, for tensors A and B, even if $\tilde{S}(A)$ and $\tilde{S}(B)$ are "small," it may still be the case that $\tilde{S}(A+B)$ is large. This third tool shows that if A has an additional property, then one can still bound $\tilde{S}(A+B)$. Roughly, the property that A must satisfy is that not only is $\tilde{S}(A)$ small, but a related notion called the "x-rank" of A must also be small.

In particular, we will remark that our three tools for bounding $\tilde{S}(T)$ strictly generalize similar tools introduced by [3] for bounding $\tilde{I}(T)$ (the *asymptotic independence number* of T, which is a weaker notion than $\tilde{S}(T)$; it is sometimes also called the "galactic subrank" or the "monomial degeneration subrank"). Hence, we generalize all of their results bounding $\omega_g(T)$ for various tensors T to at least as strong a bound on $\omega_u(T)$. See Section 1.3 below for more about prior work on bounding $\tilde{S}$. It is less clear how to compare our tools with this prior work beyond what we mention there, as prior work typically focuses on different tensors or different fields; our bounds hold over *every* field.

Section 4: Universal Method lower bounds. We apply our tools to prove upper bounds on $\tilde{S}(T)$, and hence lower bounds on $\omega_u(T)$, for a number of tensors T of interest. To prove Theorem 1.3, we show that *all three* tools can be applied to CW_q. We also apply our tools to many other tensors of interest including the generalized Coppersmith–Winograd tensors $CW_{q,\sigma}$, the generalized small Coppersmith–Winograd tensors $cw_{q,\sigma}$, the structural tensor T_q of the cyclic group C_q as well as its "lower triangular version" T_q^{lower} , and the subtensor t_{112} of $CW_q^{\otimes 2}$ which arises in [19, 21, 36, 26, 4, 25, 27]. Throughout Section 4 we give many tables of concrete lower bounds that we prove for the tensors in all these different families.

Section 5: "Completeness" of the Laser Method. Finally, we study the Laser Method. The Laser Method applied to a tensor T shows that powers $T^{\otimes n}$ can zero out into large matrix multiplication tensors. Using the properties of $\tilde{\mathrm{S}}$ that we prove in Section 2.6, we will show that the Laser Method can also be applied to a tensor T to prove a lower bound on $\tilde{\mathrm{S}}(T)$.

We prove Theorem 1.4 by combining this construction with Theorem 3.4, one of our upper-bound tools $\tilde{\mathrm{S}}(T)$. Intuitively, both Theorem 3.4 and the Laser Method are concerned with probability distributions on blocks of a tensor, and both involve counting the number of variables in powers $T^{\otimes n}$ that are consistent with these distributions. We use this intuition to show that the upper bound given by Theorem 3.4 is equal to the lower bound given by the Laser Method.

This also implies that for any laser-ready tensor T, including CW_q, cw_q, T_q, and all the other tensors we study in Section 4, our tools are tight, meaning they not only give an upper bound on $\tilde{\mathrm{S}}(T)$, but they also give a matching lower bound. Hence, for these tensors T, no better lower bound on $\omega_u(T)$ is possible by arguing only about $\tilde{\mathrm{S}}(T)$.

Our proof of Theorem 1.4 also sheds light on a notion related to the asymptotic slice rank $\tilde{\mathrm{S}}(T)$ of a tensor T, called the *asymptotic subrank* $\tilde{\mathrm{Q}}(T)$ of T. $\tilde{\mathrm{Q}}$ is a "dual" notion of asymptotic rank, and it is important in the definition of Strassen's asymptotic spectrum of tensors [32].

It is not hard to see (and follows, for instance, from Proposition 2.3 and Proposition 2.4 below) that $\tilde{\mathrm{Q}}(T) \leq \tilde{\mathrm{S}}(T)$ for all tensors T. However, there are no known separations between the two notions; whether there exists a tensor T such that $\tilde{\mathrm{Q}}(T) < \tilde{\mathrm{S}}(T)$ is an open question. As a corollary of Theorem 1.4, we prove:

Corollary 1.5. *Every laser-ready tensor T has $\tilde{\mathrm{Q}}(T) = \tilde{\mathrm{S}}(T)$.*

Since, as discussed above, almost all of the most-studied tensors are laser-ready, this might help explain why we have been unable to separate the two notions.

1.3 Other related work

Probabilistic tensors and support rank. Cohn and Umans [17] introduced the notion of the *support rank* of tensors, and showed that upper bounds on the support rank of matrix multiplication tensors can be used to design faster *Boolean* matrix multiplication algorithms. Recently, Karppa and Kaski [23] used "probabilistic tensors" as another way to design Boolean matrix multiplication algorithms.

In fact, our tools for proving asymptotic slice rank upper bounds can be used to prove lower bounds on these approaches as well. For instance, our results imply that finding a "weighted" matrix multiplication tensor as a degeneration of a power of CW_q (in order to prove a support rank upper bound) cannot result in a better exponent for Boolean matrix multiplication than 2.16805.

This is because "weighted" matrix multiplication tensors can degenerate into independent tensors just as large as their unweighted counterparts. Similarly, if a probabilistic tensor $\mathcal{T}$ is degenerated into a (probabilistic) matrix multiplication tensor, Karppa and Kaski show that this gives a corresponding support rank expression for matrix multiplication as well, and so upper bounds on $\tilde{S}(T)$ for any T in the support of $\mathcal{T}$ also result in lower bounds on this approach.

Rectangular matrix multiplication. Our tools can also be used to prove lower bounds on approaches to designing rectangular matrix multiplication algorithms. For instance, the best known rectangular matrix multiplication algorithms [27] show that powers of CW_q zero out into large rectangular matrix multiplication tensors. Using the fact that CW_q is *variable-symmetric*, this implies a corresponding upper bound on $\omega_u(CW_q)$,[3] against which our tools give a lower bound.

Slice Rank upper bounds. Our limitation results critically make use of upper bounds on the *asymptotic slice rank* of CW_q and other tensors of interest. Slice rank was first introduced by Tao [34] in a symmetric formulation of the recent proof of the capset bound [20, 22], which shows how to prove slice rank upper bounds using the "polynomial method." Since then, a number of papers have focused on proving slice rank upper bounds for many different tensors. Sawin and Tao [35, Proposition 6] show slice rank upper bounds by studying the combinatorics of the support of the power of a fixed tensor, and Naslund and Sawin [28] use that approach to study sunflower-free sets;[4] one of our slice rank bounding tools, Theorem 3.4, uses this type of approach applied to blocked tensors. Slice rank was first used in the context of matrix multiplication by Blasiak et al. [9], and this line of work has led to more techniques for proving slice rank upper bounds, including connections to the notion of instability from geometric invariant theory [9], and a generalization of the polynomial method to the nonabelian setting [10]. Christandl et al. [12] show how to characterize the asymptotic slice rank of tensors over $\mathbb{C}$ as the minimum value of a "quantum functional."

Concurrent work. Christandl, Vrana and Zuiddam [13] independently proved some of the same lower bounds on ω_u as us, including Theorem 1.3. Although we achieve the same upper bounds on $\omega_u(T)$ for a number of tensors, our techniques seem different: we use simple combinatorial tools generalizing those from our prior work [3], while their bounds use the seemingly more complicated machinery of Strassen's asymptotic spectrum of tensors [33], showing how Strassen's powerful theory comes into play in this setting. They thus phrase their results in terms of the asymptotic subrank $\tilde{Q}(T)$ of tensors rather than the asymptotic slice rank $\tilde{S}(T)$, and the fact that their bounds are often the same as ours is related to the fact we prove, in Corollary 1.5, that $\tilde{Q}(T) = \tilde{S}(T)$ for all of the tensors we study; see the bottom of Section 2.6 for a more technical discussion of the differences between the two notions. Our other results and applications of our techniques are, as far as we know, entirely new, including our matching lower bounds for $\tilde{S}(CW_q)$, $\tilde{S}(cw_q)$, and $\tilde{S}(T_q)$, bounding the value $V_\tau(T)$ of tensors, and all our results about the completeness of the Laser Method.

---

[3]If $CW_q^{\otimes n}$ degenerates to $\langle a,b,c\rangle$, then by symmetry, $CW_q^{\otimes 3n}$ degenerates to $\langle abc,abc,abc\rangle$, yielding $\omega_u(CW_q) \leq 3n\log(\tilde{R}(CW_q))/\log(abc)$.

[4]In fact, the tensor T whose slice rank is bounded in [28, Section 3] can be viewed as a change of basis of a Generalized Simple Coppersmith–Winograd tensor $cw_{q,\sigma}$ which we study below in Section 4.2.

2 Preliminaries

We begin by introducing the relevant notions and notation related to tensors and matrix multiplication. We use the same notions as in the standard literature (e. g., [31, 8, 11]), but using the notation from past work on the Galactic and Universal Methods. In particular, we will use the same notation introduced in [3, Section 3], and readers familiar with that paper may skip to Subsection 2.5.

2.1 Tensor basics

For sets $X = \{x_1, \ldots, x_q\}$, $Y = \{y_1, \ldots, y_r\}$, and $Z = \{z_1, \ldots, z_s\}$ of formal variables, a *tensor over* X,Y,Z is a trilinear form

$$T = \sum_{x_i \in X, y_j \in Y, z_k \in Z} \alpha_{ijk} x_i y_j z_k,$$

where the α_{ijk} coefficients come from an underlying field $\mathbb{F}$. The *terms*, which we write as $x_i y_j z_k$, are sometimes written as $x_i \otimes y_j \otimes z_k$ in the literature. We say T is *minimal for* X,Y,Z if, for each $x_i \in X$, there is a term involving x_i with a nonzero coefficient in T, and similarly for Y and Z (i. e., T can't be seen as a tensor over a strict subset of the variables). We say that two tensors T_1, T_2 are *isomorphic*, written $T_1 \simeq T_2$, if they are equal up to renaming variables.

If T_1 is a tensor over X_1,Y_1,Z_1, and T_2 is a tensor over X_2,Y_2,Z_2, then the *tensor product* $T_1 \otimes T_2$ is a tensor over $X_1 \times X_2, Y_1 \times Y_2, Z_1 \times Z_2$ such that, for any $(x_1,x_2) \in X_1 \times X_2$, $(y_1,y_2) \in Y_1 \times Y_2$, and $(z_1,z_2) \in Z_1 \times Z_2$, the coefficient of $(x_1,x_2)(y_1,y_2)(z_1,z_2)$ in $T_1 \otimes T_2$ is the product of the coefficient of $x_1y_1z_1$ in T_1, and the coefficient of $x_2y_2z_2$ in T_2. For any tensor T and positive integer n, the tensor power $T^{\otimes n}$ is the tensor over X^n, Y^n, Z^n resulting from taking the tensor product of n copies of T.

If T_1 is a tensor over X_1,Y_1,Z_1, and T_2 is a tensor over X_2,Y_2,Z_2, then the *direct sum* $T_1 \oplus T_2$ is a tensor over $X_1 \sqcup X_2$, $Y_1 \sqcup Y_2$, $Z_1 \sqcup Z_2$ which results from forcing the sets of variables to be disjoint (as in a normal disjoint union) and then summing the two tensors. For a nonnegative integer m and tensor T we write $m \odot T$ for the disjoint sum of m copies of T.

2.2 Tensor rank

A tensor T has *rank one* if there are values $a_i \in \mathbb{F}$ for each $x_i \in X$, $b_j \in \mathbb{F}$ for each $y_j \in Y$, and $c_k \in \mathbb{F}$ for each $z_k \in Z$, such that the coefficient of $x_i y_j z_k$ in T is $a_i b_j c_k$, or in other words,

$$T = \sum_{x_i \in X, y_j \in Y, z_k \in Z} a_i b_j c_k \cdot x_i y_j z_k = \left(\sum_{x_i \in X} a_i x_i\right)\left(\sum_{y_j \in Y} b_j y_j\right)\left(\sum_{z_k \in Z} c_k z_k\right).$$

The *rank* of a tensor T, denoted $\mathrm{R}(T)$, is the smallest number of rank one tensors whose sum (summing the coefficient of each term individually) is T. It is not hard to see that for tensors T and positive integers n, we always have $\mathrm{R}(T^{\otimes n}) \le \mathrm{R}(T)^n$, but for some tensors T of interest this inequality is not tight. We thus define the *asymptotic rank* of tensor T as $\tilde{\mathrm{R}}(T) := \inf_{n \in \mathbb{N}} (\mathrm{R}(T^{\otimes n}))^{1/n}$.

2.3 Matrix multiplication tensors

For positive integers a,b,c, the matrix multiplication tensor $\langle a,b,c\rangle$ is a tensor over $\{x_{ij}\}_{i\in[a],j\in[b]}$, $\{y_{jk}\}_{j\in[b],k\in[c]}$, $\{z_{ki}\}_{k\in[c],i\in[a]}$ given by

$$\langle a,b,c\rangle = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} x_{ij}y_{jk}z_{ki}.$$

It is not hard to verify that for positive integers a_1,a_2,b_1,b_2,c_1,c_2, we have $\langle a_1,b_1,c_1\rangle \otimes \langle a_2,b_2,c_2\rangle \simeq \langle a_1a_2,b_1b_2,c_1c_2\rangle$. The *exponent of matrix multiplication*, denoted ω, is defined as

$$\omega := \inf_{a,b,c\in\mathbb{N}} 3\log_{abc}(\mathrm{R}(\langle a,b,c\rangle)).$$

Because of the tensor product property above, we can alternatively define ω in a number of ways:

$$\omega = \inf_{a,b,c\in\mathbb{N}} 3\log_{abc}(\tilde{\mathrm{R}}(\langle a,b,c\rangle)) = \inf_{n\in\mathbb{N}} \log_n \tilde{\mathrm{R}}(\langle n,n,n\rangle) = \log_2(\tilde{\mathrm{R}}(\langle 2,2,2\rangle)).$$

For instance, Strassen [30] showed that $\mathrm{R}(\langle 2,2,2\rangle) \le 7$, which implies that $\omega \le \log_2(7)$.

2.4 Degenerations and the Universal Method

We now describe a very general way to transform from a tensor T_1 over X_1,Y_1,Z_1 to a tensor T_2 over X_2,Y_2,Z_2. For a formal variable λ, pick maps $\alpha : X_1\times X_2 \to \mathbb{F}[\lambda]$, $\beta : Y_1\times Y_2 \to \mathbb{F}[\lambda]$, and $\gamma : Z_1\times Z_2 \to \mathbb{F}[\lambda]$, which map pairs of variables to polynomials in λ, and pick a nonnegative integer h. Then, when you replace each $x\in X_1$ with $\sum_{x'\in X_2}\alpha(x,x')x'$, each $y\in Y_1$ with $\sum_{y'\in Y_2}\beta(y,y')y'$, and each $z\in Z_1$ with $\sum_{z'\in Z_2}\gamma(z,z')z'$, in T_1, then the resulting tensor T' is a tensor over X_2,Y_2,Z_2 with coefficients over $\mathbb{F}[\lambda]$. When T' is instead viewed as a polynomial in λ whose coefficients are tensors over X_2,Y_2,Z_2 with coefficients in $\mathbb{F}$, it must be that T_2 is the coefficient of λ^h, and the coefficient of $\lambda^{h'}$ is 0 for all $h' < h$.

If such a transformation is possible, we say T_2 is a *degeneration* of T_1. There are also two more restrictive types of degenerations:

- T_2 is a *monomial degeneration* of T_1 if such a transformation is possible where the polynomials in the ranges of α,β,γ have at most one monomial, and furthermore, for each $x\in X_1$ or $x'\in X_2$, there is at most one $x'\in X_2$ or $x\in X_1$, respectively, such that $\alpha(x,x')\neq 0$, and similarly for β and γ.[5]

- T_2 is a *zeroing out* of T_1 if, in addition to the restrictions of a monomial degeneration, the ranges of α,β,γ must be $\{0,1\}$.

[5] Some definitions of monomial degenerations, including the original by Strassen [32], do not have this second condition, or equivalently, consider a monomial degeneration to be a “restriction” composed with what we defined here. The distinction is not important for this paper, but we give this definition since it captures Strassen’s monomial degeneration from matrix multiplication tensors to independent tensors [31] (see also Proposition 2.5 below), and it is the notion against which the earlier paper [3] proved lower bounds.

Degenerations are useful in the context of matrix multiplication algorithms because degenerations cannot increase the asymptotic rank of a tensor. In other words, if T_2 is a degeneration of T_1, then $\tilde{\mathrm{R}}(T_2) \leq \tilde{\mathrm{R}}(T_1)$ [7]. It is often hard to bound the rank of matrix multiplication tensors directly, so all known approaches proceed by bounding the rank of a different tensor T and then showing that powers of T degenerate into matrix multiplication tensors.

More precisely, all known approaches fall within the following method, which we call the *Universal Method* [3] applied to a tensor T of asymptotic rank $R = \tilde{\mathrm{R}}(T)$: Consider all positive integers n, and all ways to degenerate $T^{\otimes n}$ into a disjoint sum $\bigoplus_{i=1}^{m} \langle a_i, b_i, c_i \rangle$ of matrix multiplication tensors, resulting in an upper bound on ω by the asymptotic sum inequality [29] of $\sum_{i=1}^{m} (a_i b_i c_i)^{\omega/3} \leq R^n$. Then, $\omega_u(T)$, the bound on ω from the Universal Method applied to T, is the inf over all such n and degenerations, of the resulting upper bound on ω.

In [3], two weaker versions of the Universal Method are also defined: *the Galactic Method*, in which the degeneration must be a monomial degeneration, resulting in a bound $\omega_g(T)$, and *the Solar Method*, in which the degeneration must be a zeroing out, resulting in a bound $\omega_s(T)$. To be clear, all three of these methods are very general, and we don't know the values of $\omega_s(T)$, $\omega_g(T)$, or $\omega_u(T)$ for almost any nontrivial tensors T. In fact, all the known approaches to bounding ω proceed by giving upper bounds on $\omega_s(T)$ for some carefully chosen tensors T; the most successful has been the Coppersmith–Winograd family of tensors $T = CW_q$, which has yielded all the best known bounds on ω since the 80s [18, 21, 36, 26, 4]. Indeed, the two most successful approaches, the Laser Method [32] and the Group-Theoretic Approach [16] ultimately use zeroing outs of tensors. We refer the reader to [3, Sections 3.3 and 3.4] for more details on these approaches and how they relate to the notions used here.

2.5 Tensor value

Coppersmith and Winograd [19] defined the *value* of a tensor in their analysis of the CW_q tensor. For a tensor T, and any $\tau \in [2/3, 1]$, the τ-value of T, denoted $V_\tau(T)$, is defined as follows: Consider all positive integers n, and all ways σ to degenerate $T^{\otimes n}$ into a direct sum $\bigoplus_{i=1}^{q(\sigma)} \langle a_i^\sigma, b_i^\sigma, c_i^\sigma \rangle$ of matrix multiplication tensors. Then, $V_\tau(T)$ is given by

$$V_\tau(T) := \sup_{n,\sigma} \left(\sum_{i=1}^{q(\sigma)} (a_i^\sigma b_i^\sigma c_i^\sigma)^\tau \right)^{1/n}.$$

We can then equivalently define $\omega_u(T)$ as the inf of ω_u, over all $\omega_u \in [2,3]$ such that $V_{\omega_u/3}(T) \geq \tilde{\mathrm{R}}(T)$. We can see from the power mean inequality that $V_\tau(T) \geq V_{2/3}(T)^{3\tau/2}$ for all $\tau \in [2/3, 1]$, although this bound is often not tight as there can be better degenerations of $T^{\otimes n}$ depending on the value of τ.

2.6 Asymptotic slice rank

The main new notions we will need in this paper relate to the slice rank of tensors. We say a tensor T over X, Y, Z has *x-rank* 1 if it is of the form

$$T = \left(\sum_{x \in X} \alpha_x \cdot x \right) \otimes \left(\sum_{y \in Y} \sum_{z \in Z} \beta_{y,z} \cdot y \otimes z \right) = \sum_{x \in X, y \in Y, z \in Z} \alpha_x \beta_{y,z} \cdot xyz$$

for some choices of the α and β coefficients over the base field. More generally, the x-rank of T, denoted $\mathrm{S_x}(T)$, is the minimum number of tensors of x-rank 1 whose sum is T. We can similarly define the y-rank, $\mathrm{S_y}$, and the z-rank, $\mathrm{S_z}$. Then, the *slice rank* of T, denoted $\mathrm{S}(T)$, is the minimum k such that there are tensors T_X, T_Y and T_Z with $T = T_X + T_Y + T_Z$ and $\mathrm{S_x}(T_X) + \mathrm{S_y}(T_Y) + \mathrm{S_z}(T_Z) = k$.

Unlike tensor rank, the slice-rank is not submultiplicative in general, i. e., there are tensors A and B such that $\mathrm{S}(A \otimes B) > \mathrm{S}(A) \cdot \mathrm{S}(B)$. For instance, it is not hard to see that $\mathrm{S}(CW_5) = 3$, but since it is known [36, 26, 4] that $\omega_s(CW_5) \leq 2.373$, it follows (e. g., from Theorem 2.9 below) that $\mathrm{S}(CW_q^{\otimes n}) \geq 7^{n \cdot 2/2.373 - o(n)} \geq 5.15^{n-o(n)}$. We are thus interested in the *asymptotic slice rank*, $\tilde{\mathrm{S}}(T)$, of tensors T, defined as

$$\tilde{\mathrm{S}}(T) := \sup_{n \in \mathbb{N}} [\mathrm{S}(T^{\otimes n})]^{1/n}.$$

We note a few simple properties of slice rank which will be helpful in our proofs:

Lemma 2.1. *For tensors A and B:*

(1) $\mathrm{S}(A) \leq \mathrm{S_x}(A) \leq \mathrm{R}(A)$*,*

(2) $\mathrm{S_x}(A \otimes B) \leq \mathrm{S_x}(A) \cdot \mathrm{S_x}(B)$*,*

(3) $\mathrm{S}(A+B) \leq \mathrm{S}(A) + \mathrm{S}(B)$*, and* $\mathrm{S_x}(A+B) \leq \mathrm{S_x}(A) + \mathrm{S_x}(B)$*,*

(4) $\mathrm{S}(A \otimes B) \leq \mathrm{S}(A) \cdot \max\{\mathrm{S_x}(B), \mathrm{S_y}(B), \mathrm{S_z}(B)\}$*, and*

(5) if A is a tensor over X, Y, Z, then $\mathrm{S_x}(A) \leq |X|$ *and hence* $\mathrm{S}(A) \leq \min\{|X|, |Y|, |Z|\}$.

Proof. (1) and (2) are straightforward. (3) follows since the sum of the slice rank (resp. x-rank) expressions for A and for B gives a slice rank (resp. x-rank) expression for $A+B$. To prove (4), let $m = \max\{\mathrm{S_x}(B), \mathrm{S_y}(B), \mathrm{S_z}(B)\}$, and note that if $A = A_X + A_Y + A_Z$ such that $\mathrm{S_x}(A_X) + \mathrm{S_y}(A_Y) + \mathrm{S_z}(A_Z) = \mathrm{S}(A)$, then

$$A \otimes B = A_X \otimes B + A_Y \otimes B + A_Z \otimes B,$$

and so

$$\begin{aligned} \mathrm{S}(A \otimes B) &\leq \mathrm{S}(A_X \otimes B) + \mathrm{S}(A_Y \otimes B) + \mathrm{S}(A_Z \otimes B) \\ &\leq \mathrm{S_x}(A_X \otimes B) + \mathrm{S_y}(A_Y \otimes B) + \mathrm{S_z}(A_Z \otimes B) \\ &\leq \mathrm{S_x}(A_X)\,\mathrm{S_x}(B) + \mathrm{S_y}(A_Y)\,\mathrm{S_y}(B) + \mathrm{S_z}(A_Z)\,\mathrm{S_z}(B) \\ &\leq \mathrm{S_x}(A_X)m + \mathrm{S_y}(A_Y)m + \mathrm{S_z}(A_Z)m = \mathrm{S}(A) \cdot m. \end{aligned}$$

Finally, (5) follows since, for instance, any tensor with one only x-variable has x-rank 1. □

Asymptotic slice rank is interesting in the context of matrix multiplication algorithms because of the following facts.

Definition 2.2. For a positive integer q, the *independent tensor of size* q, denoted $\langle q \rangle$, is the tensor $\sum_{i=1}^{q} x_i y_i z_i$ with q terms that do not share any variables.

Proposition 2.3 ([35, Corollary 2]). *If A and B are tensors such that A has a degeneration to B, then $\mathrm{S}(B) \le \mathrm{S}(A)$, and hence $\tilde{\mathrm{S}}(B) \le \tilde{\mathrm{S}}(A)$.*

Proposition 2.4 ([34, Lemma 1]; see also [9, Lemma 4.7]). *For any positive integer q, we have $\mathrm{S}(\langle q \rangle) = \tilde{\mathrm{S}}(\langle q \rangle) = q$, where $\langle q \rangle$ is the independent tensor of size q.*

Proposition 2.5 ([32, Theorem 6.6]; see also [3, Lemma 4.2]). *For any positive integers a,b,c, the matrix multiplication tensor $\langle a,b,c \rangle$ has a (monomial) degeneration to an independent tensor of size at least $0.75 \cdot abc/\max\{a,b,c\}$.*

Corollary 2.6. *For any positive integers a,b,c, we have $\tilde{\mathrm{S}}(\langle a,b,c \rangle) = abc/\max\{a,b,c\}$.*

Proof. Assume without loss of generality that $c \ge a,b$. For any positive integer n, we have that $\langle a,b,c \rangle^{\otimes n} \simeq \langle a^n, b^n, c^n \rangle$ has a degeneration to an independent tensor of size at least $0.75 \cdot a^n b^n$, meaning $\mathrm{S}(\langle a,b,c \rangle^{\otimes n}) \ge 0.75 \cdot a^n b^n$ and hence $\tilde{\mathrm{S}}(\langle a,b,c \rangle) \ge (0.75)^{1/n} ab$, which means $\tilde{\mathrm{S}}(\langle a,b,c \rangle) \ge ab$. Meanwhile, $\langle a,b,c \rangle$ has ab different x-variables, so it must have $\mathrm{S}_{\mathrm{x}}(\langle a,b,c \rangle) \le ab$ and more generally, $\mathrm{S}(\langle a,b,c \rangle^{\otimes n}) \le \mathrm{S}_{\mathrm{x}}(\langle a,b,c \rangle^{\otimes n}) \le (ab)^n$, which means $\tilde{\mathrm{S}}(\langle a,b,c \rangle) \le ab$. □

To summarize: we know that degenerations cannot increase asymptotic slice rank, and that matrix multiplication tensors have a high asymptotic slice rank. Hence, if T is a tensor such that $\omega_u(T)$ is "small," meaning a power of T has a degeneration to a disjoint sum of many large matrix multiplication tensors, then T itself must have "large" asymptotic slice rank. This can be formalized identically to [3, Theorem 4.1 and Corollary 4.3] to show the following.

Theorem 2.7. *For any tensor T,*

$$\tilde{\mathrm{S}}(T) \ge \tilde{\mathrm{R}}(T)^{\frac{6}{\omega_u(T)} - 2}.$$

Corollary 2.8. *For any tensor T, if $\omega_u(T) = 2$, then $\tilde{\mathrm{S}}(T) = \tilde{\mathrm{R}}(T)$. Moreover, for every constant $s < 1$, there is a constant $w > 2$ such that every tensor T with $\tilde{\mathrm{S}}(T) \le \tilde{\mathrm{R}}(T)^s$ must have $\omega_u(T) \ge w$.*

Almost all the tensors we consider in this note are *variable-symmetric* tensors, and for these tensors T we can get a better lower bound on $\omega_u(T)$ from an upper bound on $\tilde{\mathrm{S}}(T)$. We say that a tensor T over X,Y,Z is variable-symmetric if $|X| = |Y| = |Z|$, and the coefficient of $x_i y_j z_k$ equals the coefficient of $x_j y_k z_i$ in T for all $(x_i, y_j, z_k) \in X \times Y \times Z$.

Theorem 2.9. *For a variable-symmetric tensor T we have $\omega_u(T) \ge 2\log(\tilde{\mathrm{R}}(T))/\log(\tilde{\mathrm{S}}(T))$.*

Proof. As in the proof of [3, Theorem 4.1], by definition of ω_u, we know that for every $\delta > 0$, there is a positive integer n such that $T^{\otimes n}$ has a degeneration to $F \odot \langle a,b,c \rangle$ for integers F,a,b,c such that $\omega_u(T)^{1+\delta} \ge 3\log(\tilde{\mathrm{R}}(T)^n/F)/\log(abc)$. In fact, since T is symmetric, we know $T^{\otimes n}$ also has a degeneration to $F \odot \langle b,c,a \rangle$ and to $F \odot \langle c,a,b \rangle$, and so $T^{\otimes 3n}$ has a degeneration to $F^3 \odot \langle abc, abc, abc \rangle$. As above, it follows that $\tilde{\mathrm{S}}(T^{\otimes 3n}) \ge \tilde{\mathrm{S}}(F^3 \odot \langle abc, abc, abc \rangle) = F^3 \cdot (abc)^2$. Rearranging, we see

$$abc \le \tilde{\mathrm{S}}(T)^{3n/2}/F^{3/2}.$$

Hence,

$$\omega_u(T)^{1+\delta} \ge 3\frac{\log(\tilde{\mathrm{R}}(T)^n/F)}{\log(abc)} \ge 3\frac{\log(\tilde{\mathrm{R}}(T)^n/F)}{\log(\tilde{\mathrm{S}}(T)^{3n/2}/F^{3/2})} = 2\frac{\log(\tilde{\mathrm{R}}(T)) - \frac{1}{n}\log(F)}{\log(\tilde{\mathrm{S}}(T)) - \frac{1}{n}\log(F)} \ge 2\frac{\log(\tilde{\mathrm{R}}(T))}{\log(\tilde{\mathrm{S}}(T))},$$

where the last step follows because $\tilde{\mathrm{R}}(T) \geq \tilde{\mathrm{S}}(T)$ and so subtracting the same quantity from both the numerator and denominator cannot decrease the value of the fraction. This holds for all $\delta > 0$ and hence implies the desired result. $\square$

Slice rank versus subrank. For a tensor T, let $Q'(T)$ denote the largest integer q such that there is a degeneration from T to $\langle q \rangle$. The *asymptotic subrank* of T is defined as $\tilde{\mathrm{Q}}(T) := \sup_{n\in\mathbb{N}} Q'(T^{\otimes n})^{1/n}$. Proposition 2.3 and Proposition 2.4 above imply that $\tilde{\mathrm{Q}}(T) \leq \tilde{\mathrm{S}}(T)$ for all tensors T. Similarly, it is not hard to see that Theorem 2.7 and Theorem 2.9 hold with $\tilde{\mathrm{S}}$ replaced by $\tilde{\mathrm{Q}}$. One could thus conceivably hope to prove stronger lower bounds than those in this paper by bounding $\tilde{\mathrm{Q}}$ instead of $\tilde{\mathrm{S}}$. However, we will prove in Corollary 5.5 below that $\tilde{\mathrm{Q}}(T) = \tilde{\mathrm{S}}(T)$ for every tensor we study in this paper, so such an improvement using $\tilde{\mathrm{Q}}$ is impossible. More generally, there are currently no known tensors T for which the best known upper bound on $\tilde{\mathrm{Q}}(T)$ is smaller than the best known upper bound on $\tilde{\mathrm{S}}(T)$ (including the new bounds of [12, 13]; more precisely, [12] give tight upper bounds on asymptotic slice rank). Hence, new upper-bound tools for $\tilde{\mathrm{Q}}$ would be required for such an approach to proving better lower bounds on ω_u.

2.7 Partition notation

In several of our results, we will be partitioning the terms of tensors into blocks defined by partitions of the three sets of variables. Here we introduce some notation for some properties of such partitions; these definitions all depend on the particular partition of the variables being used, which will be clear from context.

Suppose T is a tensor minimal over X, Y, Z, and let $X = X_1 \cup \cdots \cup X_{k_X}$, $Y = Y_1 \cup \cdots \cup Y_{k_Y}$, $Z = Z_1 \cup \cdots \cup Z_{k_Z}$ be partitions of the three sets of variables. For $(i, j, k) \in [k_X] \times [k_Y] \times [k_Z]$, let T_{ijk} be T restricted to X_i, Y_j, Z_k (i. e., T with $X \setminus X_i$, $Y \setminus Y_j$, and $Z \setminus Z_k$ zeroed out), and let

$$L = \{T_{ijk} \mid (i, j, k) \in [k_X] \times [k_Y] \times [k_Z], T_{ijk} \neq 0\}.$$

T_{ijk} is called a *block* of T. For $i \in [k_X]$ let $L_{X_i} = \{T_{ij'k'} \in L \mid (j', k') \in [k_Y] \times [k_Z]\}$, and define similarly L_{Y_j} and L_{Z_k}.

We will be particularly interested in probability distributions $p : L \to [0, 1]$. Let $P(L)$ be the set of such distributions. For such a $p \in P(L)$, and for $i \in [k_X]$, let $p(X_i) := \sum_{T_{ijk} \in L_{X_i}} p(T_{ijk})$, and similarly $p(Y_j)$ and $p(Z_k)$. Then, define $p_X \in \mathbb{R}$ by

$$p_X := \prod_{i\in[k_X]} \left(\frac{|X_i|}{p(X_i)} \right)^{p(X_i)},$$

and p_Y and p_Z similarly. This expression, which arises naturally in the Laser Method, and is related to the Shannon entropy of p, will play an important role in our upper bounds and lower bounds.

2.8 Tensor rotations and variable-symmetric tensors

If T is a tensor over $X = \{x_1, \ldots, x_q\}$, $Y = \{y_1, \ldots, y_r\}$, and $Z = \{z_1, \ldots, z_s\}$, then the *rotation of* T, denoted *rot*(T), is the tensor over $\{x_1, \ldots, x_r\}$, $\{y_1, \ldots, y_s\}$, and $\{z_1, \ldots, z_q\}$ such that for any $(x_i, y_j, z_k) \in$

$X \times Y \times Z$, the coefficient of $x_i y_j z_k$ in T is equal to the coefficient of $x_j y_k z_i$ in $rot(T)$. Tensor T is *variable-symmetric* if $T \simeq rot(T)$.

If T is a variable-symmetric tensor minimal over X, Y, Z, then partitions $X = X_1 \cup \cdots \cup X_{k_X}$, $Y = Y_1 \cup \cdots \cup Y_{k_Y}$, $Z = Z_1 \cup \cdots \cup Z_{k_Z}$ of the sets of variables are called T*-symmetric* if (using the notation of the previous subsection) $k_X = k_Y = k_Z$, $|X_i| = |Y_i| = |Z_i|$ for all $i \in [k_X]$, and the block $T_{jki} \simeq rot(T_{ijk})$ for all $(i, j, k) \in [k_X]^3$. For the L resulting from such a T-symmetric partition, a probability distribution $p \in P(L)$ is called T*-symmetric* if it satisfies $p(T_{ijk}) = p(T_{jki})$ for all $(i, j, k) \in [k_X]^3$, and we write $P^{sym}(L) \subseteq P(L)$ for the set of such T-symmetric distributions. Notice in particular that any $p \in P^{sym}(L)$ satisfies $p_X = p_Y = p_Z$.

3 Combinatorial tools for asymptotic upper bounds on slice rank

We now give three general tools for proving upper bounds on $\tilde{\mathrm{S}}(T)$ for many tensors T. Each of our tools generalizes one of the three main tools of [3], which were bounding the weaker notion $\tilde{I}$ instead of $\tilde{\mathrm{S}}$, and could also only apply to a more restrictive set of tensors. We will make clear what previous result we are generalizing, although our presentation here is entirely self-contained.

3.1 Generalization of [3, Theorem 5.3]

We know from Lemma 2.1 part (5) that tensors T over X, Y, Z such that $\min\{|X|, |Y|, |Z|\}$ is small must also have small $\tilde{\mathrm{S}}(T)$. We begin here by showing that if T can be written as a sum of a few tensors, each of which has a small bound on $\tilde{\mathrm{S}}$ which can be proved in this way, then we can still prove an upper bound on $\tilde{\mathrm{S}}(T)$.

The *measure* of a tensor T, denoted $\mu(T)$, is given by $\mu(T) := |X| \cdot |Y| \cdot |Z|$, where X, Y, Z are the sets of variables which are minimal for T. We state two simple facts about μ:

Fact 3.1. *For tensors A and B,*

- $\mu(A \otimes B) = \mu(A) \cdot \mu(B)$*, and*
- *if A is minimal over* X, Y, Z*, then* $\mathrm{S}(A) \leq \min\{|X|, |Y|, |Z|\} \leq \mu(A)^{1/3}$.

Theorem 3.2. *Suppose T is a tensor, and $T_1, \ldots, T_k$ are tensors with $T = T_1 + \cdots + T_k$. Then,* $\tilde{\mathrm{S}}(T) \leq \sum_{i=1}^{k} (\mu(T_i))^{1/3}$.

Proof. Note that

$$T^{\otimes n} = \sum_{(P_1, \ldots, P_n) \in \{T_1, \ldots, T_k\}^n} P_1 \otimes \cdots \otimes P_n.$$

It follows that

$$\begin{aligned}
\mathrm{S}(T^{\otimes n}) &\le \sum_{(P_1,\ldots,P_n)\in\{T_1,\ldots,T_k\}^n} \mathrm{S}(P_1\otimes\cdots\otimes P_n)\\
&\le \sum_{(P_1,\ldots,P_n)\in\{T_1,\ldots,T_k\}^n} \mu(P_1\otimes\cdots\otimes P_n)^{1/3}\\
&= \sum_{(P_1,\ldots,P_n)\in\{T_1,\ldots,T_k\}^n} (\mu(P_1)\cdot\mu(P_2)\cdots\mu(P_n))^{1/3}\\
&= (\mu(T_1)^{1/3}+\cdots+\mu(T_k)^{1/3})^n,
\end{aligned}$$

which implies as desired that $\tilde{\mathrm{S}}(T)\le \mu(T_1)^{1/3}+\cdots+\mu(T_k)^{1/3}$. □

Remark 3.3. [3, Theorem 5.3], in addition to bounding $\tilde{I}$ instead of $\tilde{\mathrm{S}}$, also required that $T = T_1+\cdots+T_k$ be a *partition* of the terms of T. Here in Theorem 3.2 we are allowed any tensor sum, although in general a partition minimizes the resulting upper bound.

3.2 Generalization of [3, Theorem 5.2]

This tool will be the most important in proving upper bounds on the asymptotic slice rank of many tensors of interest. We show that a partitioning method similar to the Laser Method applied to a tensor T can be used to prove upper bounds on $\tilde{\mathrm{S}}(T)$. Recall the definitions and notation about partitions of tensors from Section 2.7.

Theorem 3.4. *For any tensor T and partition of its sets of variables,*

$$\tilde{\mathrm{S}}(T)\le \sup_{p\in P(L)} \min\{p_X,p_Y,p_Z\}.$$

Our proof of Theorem 3.4 uses one of the most successful techniques from past work on proving slice rank upper bounds, of partitioning the terms of $T^{\otimes n}$ by distribution from $P(L)$; similar results (without blocking variables) were also proved, for instance, by [35, Proposition 6], [3, Theorem 5.2], [12, Theorem 5.9].

Proof of Theorem 3.4. For any positive integer n, we can write

$$T^{\otimes n} = \sum_{(P_1,\ldots,P_n)\in L^n} P_1\otimes\cdots\otimes P_n.$$

For a given $(P_1,\ldots,P_n)\in L^n$, let $dist(P_1,\ldots,P_n)$ be the probability distribution on L which results from picking a uniformly random $\alpha\in[n]$ and outputting P_α. For a probability distribution $p: L\to[0,1]$, define $L_{n,p} := \{(P_1,\ldots,P_n)\in L^n \mid dist(P_1,\ldots,P_n)=p\}$. Note that the number of p for which $L_{n,p}$ is nonempty is only $\mathrm{poly}(n)$, since they are the distributions which assign an integer multiple of $1/n$ to each element of L. Let D be the set of these probability distributions.

We can now rearrange:

$$T^{\otimes n} = \sum_{p\in D}\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n.$$

Hence,

$$\begin{aligned}\mathrm{S}(T^{\otimes n}) &\leq \sum_{p\in D} \mathrm{S}\left(\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n\right)\\ &\leq \mathrm{poly}(n)\cdot\max_{p\in D}\mathrm{S}\left(\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n\right).\end{aligned}$$

For any probability distribution $p : L \to [0,1]$, let us count the number of x-variables used in

$$\left(\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n\right).$$

These are the tuples of the form $(x_1,\ldots,x_n)\in X^n$ where, for each $i\in[k_X]$, there are exactly $n\cdot p(X_i)$ choices of j for which $x_j\in X_i$. The number of these is,[6]

$$\binom{n}{n\cdot p(X_1), n\cdot p(X_2),\ldots,n\cdot p(X_{k_X})}\cdot\prod_{i\in[k_X]}|X_i|^{n\cdot p(X_i)}.$$

This is at most $p_X^{n+o(n)}$, where p_X is the quantity defined in Section 2.7. It follows that

$$\mathrm{S}_{\mathrm{x}}\left(\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n\right)\leq p_X^{n+o(n)}.$$

We can similarly argue about S_{y} and S_{z}. Hence,

$$\begin{aligned}\mathrm{S}(T^{\otimes n}) &\leq \mathrm{poly}(n)\cdot\max_{p\in D}\mathrm{S}\left(\sum_{(P_1,\ldots,P_n)\in L_{n,p}} P_1\otimes\cdots\otimes P_n\right)\\ &\leq \mathrm{poly}(n)\cdot\max_{p\in D}\min\{p_X,p_Y,p_Z\}^{n+o(n)}\\ &\leq \mathrm{poly}(n)\cdot\sup_{p\in P(L)}\min\{p_X,p_Y,p_Z\}^{n+o(n)}.\end{aligned}$$

Hence, $\mathrm{S}(T^{\otimes n})\leq \sup_p \min\{p_X,p_Y,p_Z\}^{n+o(n)}$, and the desired result follows. □

[6] Here,

$$\binom{n}{p_1 n, p_2 n,\ldots,p_\ell n} = \frac{n!}{(p_1 n)!(p_2 n)!\cdots(p_\ell n)!}$$

with each $p_i\in[0,1]$ and $p_1+\cdots+p_\ell=1$, is the multinomial coefficient, with the known bound from Stirling's approximation, for fixed p_is, that

$$\binom{n}{p_1 n, p_2 n,\ldots,p_\ell n}\leq\left(\prod_i p_i^{-p_i}\right)^{n+o(n)}.$$

Throughout this paper we use the convention that $p_i^{p_i}=1$ when $p_i=0$.

Remark 3.5. [3, Theorem 5.2] is less powerful than our Theorem 3.4 in two ways: it used $\tilde{I}$ instead of $\tilde{\mathrm{S}}$, and it required each X_i, Y_j, Z_k to contain only one variable.

Remark 3.6. Suppose T is over X,Y,Z with $|X| = |Y| = |Z| = q$. For any probability distribution p we always have $p_X, p_Y, p_Z \le q$, and moreover we only have $p_X = q$ when $p(X_i) = |X_i|/q$ for each i. Similar to [3, Corollary 5.1], it follows that if no probability distribution p is δ-close (say, in ℓ_1 distance) to having $p(X_i) = |X_i|/q$ for all i, $p(Y_j) = |Y_j|/q$ for all j, and $p(Z_k) = |Z_k|/q$ for all k, simultaneously, then we get $\tilde{\mathrm{S}}(T) \le q^{1-f(\delta)}$ for some increasing function f with $f(\delta) > 0$ for all $\delta > 0$.

Remark 3.7. Applying Theorem 3.4 when all the parts of the partitions of X, Y, and Z have size 1 always yields the best possible bound that Theorem 3.4 can give on $\tilde{\mathrm{S}}(T)$. That said, many tensors are easier to argue about when a different partition is used, as we will see in Section 5 below.

Before moving on, we make an observation about applying Theorem 3.4 to variable-symmetric tensors. This observation has implicitly been used in past work on applying the Laser Method, such as [19], but we prove it here for completeness. Recall the notation in Section 2.8 about such tensors.

Proposition 3.8. *Suppose T is a variable-symmetric tensor over X,Y,Z, and $X = X_1 \cup \cdots \cup X_{k_X}$, $Y = Y_1 \cup \cdots \cup Y_{k_Y}$, $Z = Z_1 \cup \cdots \cup Z_{k_Z}$ are T-symmetric partitions. Then,*

$$\tilde{\mathrm{S}}(T) \le \sup_{p \in P^{sym}(L)} p_X.$$

Proof. We know from Theorem 3.4 that $\tilde{\mathrm{S}}(T) \le \sup_{p \in P(L)} \min\{p_X, p_Y, p_Z\}$. We will show that for any $p \in P(L)$, there is a $p' \in P^{sym}(L)$ such that $\min\{p_X, p_Y, p_Z\} \le \min\{p'_X, p'_Y, p'_Z\}$, which means that in fact, $\tilde{\mathrm{S}}(T) \le \sup_{p \in P^{sym}(L)} \min\{p_X, p_Y, p_Z\}$. Finally, the desired result will follow since, for any $p' \in P^{sym}(L)$, we have $p'_X = p'_Y = p'_Z$.

Consider any $p \in P(L)$, and define the distribution $p' \in P^{sym}(L)$ by $p'(T_{ijk}) := (p(T_{ijk}) + p(T_{jki}) + p(T_{kij}))/3$ for each $T_{ijk} \in L$. To prove that $\min\{p_X, p_Y, p_Z\} \le p'_X$, we will show that $(p_X p_Y p_Z)^{1/3} \le p'_X$.

$$\begin{aligned}
(p_X p_Y p_Z)^{1/3} &= \prod_{i \in [k_X]} \left(\frac{|X_i|}{p(X_i)}\right)^{p(X_i)/3} \left(\frac{|Y_i|}{p(Y_i)}\right)^{p(Y_i)/3} \left(\frac{|Z_i|}{p(Z_i)}\right)^{p(Z_i)/3} \\
&= \prod_{i \in [k_X]} \frac{|X_i|^{p'(X_i)}}{(p(X_i)^{p(X_i)} p(Y_i)^{p(Y_i)} p(Z_i)^{p(Z_i)})^{1/3}} \\
&\le \prod_{i \in [k_X]} \frac{|X_i|^{p'(X_i)}}{p'(X_i)^{p'(X_i)}} \\
&= p'_X,
\end{aligned}$$

where the second-to-last step follows from the fact that for any real numbers $a,b,c \in [0,1]$, setting $d = (a+b+c)/3$, we have $a^a b^b c^c \ge d^{3d}$. $\square$

3.3 Generalization of [3, Theorem 5.1]

The final remaining tool from [3], their Theorem 5.1, turns out to be unnecessary for proving our tight lower bounds in the next section. Nonetheless, we sketch here how to extend it to give asymptotic slice rank upper bounds as well.

For a tensor T, let $m(T) := \max\{\mathrm{S_x}(T), \mathrm{S_y}(T), \mathrm{S_z}(T)\}$. Recall from Lemma 2.1 that for any two tensors A, B we have $\mathrm{S}(A \otimes B) \leq \mathrm{S}(A) \cdot m(B)$.

In general, for two tensors A and B, even if $\tilde{\mathrm{S}}(A)$ and $\tilde{\mathrm{S}}(B)$ are "small," it might still be the case that $\tilde{\mathrm{S}}(A+B)$ is "large," much larger than $\tilde{\mathrm{S}}(A) + \tilde{\mathrm{S}}(B)$. For instance, for any positive integer q, define the tensors $T_1 := \sum_{i=0}^{q} x_0 y_i z_i$, $T_2 := \sum_{i=1}^{q+1} x_i y_0 z_i$, and $T_3 := \sum_{i=1}^{q+1} x_i y_i z_{q+1}$. We can see that $\tilde{\mathrm{S}}(T_1) = \tilde{\mathrm{S}}(T_2) = \tilde{\mathrm{S}}(T_3) = 1$, but $T_1 + T_2 + T_3 = CW_q$, and we will show soon that $\tilde{\mathrm{S}}(CW_q)$ grows unboundedly with q.

Here we show that if, not only is $\tilde{\mathrm{S}}(A)$ small, but even $\mathrm{S_x}(A)$ is small, then we can get a decent bound on $\tilde{\mathrm{S}}(A+B)$.

Theorem 3.9. *Suppose T, A, B are tensors such that $A + B = T$. Then,*

$$\tilde{\mathrm{S}}(T) \leq \left(\frac{m(A)}{(1-p)\cdot \mathrm{S_x}(A)}\right)^{1-p} \cdot \frac{1}{p^p},$$

where $p \in [0,1]$ is given by

$$p := \frac{\log\left(\frac{\mathrm{S_x}(B)}{\tilde{\mathrm{S}}(B)}\right)}{\log\left(\frac{m(A)}{\mathrm{S_x}(A)}\right) + \log\left(\frac{\mathrm{S_x}(B)}{\tilde{\mathrm{S}}(B)}\right)}.$$

Proof. We begin by, for any integers $n \geq k \geq 0$, giving bounds on $\mathrm{S}(A^{\otimes k} \otimes B^{\otimes(n-k)})$. First, since $\mathrm{S_x}$ is submultiplicative, we have

$$\mathrm{S}(A^{\otimes k} \otimes B^{\otimes(n-k)}) \leq \mathrm{S_x}(A^{\otimes k} \otimes B^{\otimes(n-k)}) \leq \mathrm{S_x}(A)^k \cdot \mathrm{S_x}(B)^{n-k}.$$

Second, from the definition of m, we have

$$\mathrm{S}(A^{\otimes k} \otimes B^{\otimes(n-k)}) \leq m(A^{\otimes k}) \cdot \mathrm{S}(B^{\otimes(n-k)}) \leq m(A)^k \cdot \tilde{\mathrm{S}}(B)^{n-k}.$$

It follows that for any positive integer n we have

$$\mathrm{S}(T^{\otimes n}) \leq \sum_{k=0}^{n} \binom{n}{k} \cdot \mathrm{S}(A^{\otimes k} \otimes B^{\otimes(n-k)}) \leq \sum_{k=0}^{n} \binom{n}{k} \cdot \min\{\mathrm{S_x}(A)^k \cdot \mathrm{S_x}(B)^{n-k}, m(A)^k \cdot \tilde{\mathrm{S}}(B)^{n-k}\}.$$

As in the proof of [3, Theorem 5.1], we can see that the quantity $\binom{n}{k} \cdot \min\{\mathrm{S_x}(A)^k \cdot \mathrm{S_x}(B)^{n-k}, m(A)^k \cdot \tilde{\mathrm{S}}(B)^{n-k}\}$ is maximized at $k = pn$, and the result follows. □

Remark 3.10. This result generalizes [3, Theorem 5.1], no longer requiring that A be the tensor T restricted to a single x-variable. In [3, Theorem 5.1], since A is T restricted to a single x-variable, and we required A to have at most q terms, we got the bounds $\mathrm{S_x}(A) = 1$ and $m(A) \leq q$. Similarly, B had at most $q-1$ different x-variables, so $\mathrm{S_x}(B) \leq q-1$. Substituting those values into Theorem 3.9 yields the original [3, Theorem 5.1] with $\tilde{I}$ replaced by $\tilde{\mathrm{S}}$.

4 Computing the slice ranks for tensors of interest

In this section, we give slice rank upper bounds for a number of tensors of interest. It will follow from Section 5 that *all of the bounds we prove in this section are tight*.

4.1 Generalized Coppersmith–Winograd tensors

We begin with the generalized CW tensors defined in [3], which for a positive integer q and a permutation $\sigma : [q] \to [q]$ are given by

$$CW_{q,\sigma} := x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^{q}(x_iy_{\sigma(i)}z_0 + x_iy_0z_i + x_0y_iz_i).$$

The usual Coppersmith–Winograd tensor CW_q results by setting σ to the identity permutation. Just as in [3, Section 7.1], we can see that Theorem 3.2 and Theorem 3.4 immediately apply to $CW_{q,\sigma}$ to show that there is a universal constant $\delta > 0$ such that for any q and σ we have $\tilde{\mathrm{S}}(CW_{q,\sigma}) \le (q+2)^{1-\delta}$, and hence a universal constant $c > 2$ such that $\omega_u(CW_{q,\sigma}) \ge c$. Indeed, by proceeding in this way, we get the exact same constants as in [3].

That said, we will now use Theorem 3.4 to prove that $c \ge 2.16805$. (In fact, essentially the same argument as we present now shows that [3, Theorem 5.2] was already sufficient to show the weaker claim that $\omega_g(CW_{q,\sigma}) \ge 2.16805$.)

We begin by partitioning the sets of variables of $CW_{q,\sigma}$, using the notation of Theorem 3.4. Let $X_0 = \{x_0\}$, $X_1 = \{x_1,\dots,x_q\}$, and $X_2 = \{x_{q+1}\}$, so that $X_0 \cup X_1 \cup X_2$ is a partition of the x-variables of $CW_{q,\sigma}$.[7] Similarly, let $Y_0 = \{y_0\}$, $Y_1 = \{y_1,\dots,y_q\}$, $Y_2 = \{y_{q+1}\}$, $Z_0 = \{z_0\}$, $Z_1 = \{z_1,\dots,z_q\}$, and $Z_2 = \{z_{q+1}\}$. We can see this is a $CW_{q,\sigma}$-symmetric partition with $L = \{T_{002}, T_{020}, T_{200}, T_{011}, T_{101}, T_{110}\}$.

Consider any probability distribution $p \in P^{sym}(L)$. By symmetry, we know that $p(T_{002}) = p(T_{020}) = p(T_{200}) = v$ and $p(T_{011}) = p(T_{101}) = p(T_{110}) = 1/3 - v$ for some value $v \in [0,1/3]$. Applying Theorem 3.4, and in particular Proposition 3.8, yields:

$$\tilde{\mathrm{S}}(CW_q) \le \sup_{v\in[0,1/3]} \frac{q^{2(1/3-v)}}{v^v(2/3-2v)^{2/3-2v}(1/3+v)^{1/3+v}}.$$

In fact, we will see in the next section that this is tight (i. e., the value above is *equal* to $\tilde{\mathrm{S}}(CW_q)$, not just an upper bound on it). The values for the first few q can be computed using optimization software as follows:

[7] The sets of partitions were 1-indexed before, but we 0-index here for notational consistency with past work.

q	$\tilde{\mathrm{S}}(CW_{q,\sigma})$
1	$2.7551\cdots$
2	$3.57165\cdots$
3	$4.34413\cdots$
4	$5.07744\cdots$
5	$5.77629\cdots$
6	$6.44493\cdots$
7	$7.08706\cdots$
8	$7.70581\cdots$

Finally, using the lower bound $\tilde{\mathrm{R}}(CW_{q,\sigma}) \geq q+2$ (in fact, it is known that $\tilde{\mathrm{R}}(CW_{q,\sigma}) = q+2$), and the upper bound on $\tilde{\mathrm{S}}(CW_{q,\sigma})$ we just proved, we can apply Theorem 2.9 to give lower bounds $\omega_u(CW_{q,\sigma}) \geq 2\log(\tilde{\mathrm{R}}(CW_{q,\sigma}))/\log(\tilde{\mathrm{S}}(CW_{q,\sigma})) \geq 2\log(q+2)/\log(\tilde{\mathrm{S}}(CW_{q,\sigma}))$ as follows:

q	Lower Bound on $\omega_u(CW_{q,\sigma})$
1	$2.16805\cdots$
2	$2.17794\cdots$
3	$2.19146\cdots$
4	$2.20550\cdots$
5	$2.21912\cdots$
6	$2.23200\cdots$
7	$2.24404\cdots$
8	$2.25525\cdots$

It is not hard to see that the resulting lower bound on $\omega_u(CW_{q,\sigma})$ is increasing with q and is always at least $2.16805\ldots$ (see Appendix A below for a proof), and hence that for any q and any σ we have $\omega_u(CW_{q,\sigma}) \geq 2.16805$ as desired.

4.2 Generalized simple Coppersmith–Winograd tensors

Similarly to $CW_{q,\sigma}$, we can define for a positive integer q and a permutation $\sigma : [q] \to [q]$ the simple Coppersmith–Winograd tensor $cw_{q,\sigma}$ given by:

$$cw_{q,\sigma} := \sum_{i=1}^{q} (x_i y_{\sigma(i)} z_0 + x_i y_0 z_i + x_0 y_i z_i).$$

These tensors, when σ is the identity permutation id, are well-studied. For instance, Coppersmith and Winograd [19] showed that if $\tilde{\mathrm{R}}(cw_{2,id}) = 2$ then $\omega = 2$.

We will again give a tight bound on $\tilde{\mathrm{S}}(cw_{q,\sigma})$ using Theorem 3.4 combined with the next section. To apply Theorem 3.4, and in particular Proposition 3.8, we again pick a partition of the variables. Let $X_0 = \{x_0\}$, $X_1 = \{x_1,\ldots,x_q\}$, $Y_0 = \{y_0\}$, $Y_1 = \{y_1,\ldots,y_q\}$, $Z_0 = \{z_0\}$, and $Z_1 = \{z_1,\ldots,z_q\}$. This is a $cw_{q,\sigma}$-symmetric partition with $L = \{T_{011}, T_{101}, T_{110}\}$. There is a unique $p \in P^{sym}(L)$, which assigns probability $1/3$ to each part. It follows that

$$\tilde{\mathrm{S}}(cw_{q,\sigma}) \leq (1/3)^{-1/3}(2/3)^{-2/3} \cdot q^{2/3} = \frac{3}{2^{2/3}} \cdot q^{2/3}.$$

Again, we will see in the next section that this bound is tight. Using the lower bound $\tilde{\mathrm{R}}(cw_{q,\sigma}) \geq q+1$, we get the lower bound

$$\omega_u(cw_{q,\sigma}) \geq 2\frac{\log(q+1)}{\log\left(\frac{3}{2^{2/3}} \cdot q^{2/3}\right)}.$$

The first few values are as follows; note that we cannot get a lower bound better than 2 when $q=2$ because of Coppersmith and Winograd's aforementioned remark that if $\tilde{\mathrm{R}}(cw_{2,id}) = 2$ then $\omega = 2$..

q	Lower Bound on $\omega_u(cw_{q,\sigma})$
1	$2.17795\cdots$
2	2
3	$2.02538\cdots$
4	$2.06244\cdots$
5	$2.09627\cdots$
6	$2.12549\cdots$
7	$2.15064\cdots$

4.3 Cyclic group tensors

We next look at two tensors which were studied in [16], [2], and [3, Section 7.3]. For each positive integer q, define the tensor T_q (the structural tensor of the cyclic group C_q) as:

$$T_q = \sum_{i=0}^{q-1}\sum_{j=0}^{q-1} x_i y_j z_{i+j \bmod q}.$$

Define also the lower triangular version of T_q, called T_q^{lower}, as:

$$T_q^{lower} = \sum_{i=0}^{q-1}\sum_{j=0}^{q-1-i} x_i y_j z_{i+j}.$$

While Theorem 3.4 does not give any nontrivial upper bounds on $\tilde{\mathrm{S}}(T_q)$, it does give nontrivial upper bounds on $\tilde{\mathrm{S}}(T_q^{lower})$, as noted in [3, Section 7.3]. Using computer optimization software, we can compute our lower bound on $\tilde{\mathrm{S}}(T_q^{lower})$, using Theorem 3.4 where each partition contains exactly one variable, for the first few values of q:

q	Upper Bound on $\tilde{\mathrm{S}}(T_q^{lower})$
2	$1.88988\cdots$
3	$2.75510\cdots$
4	$3.61071\cdots$
5	$4.46157\cdots$

We show in the next section that these numbers are also tight. The value of $\tilde{\mathrm{S}}(T_q^{lower})$ also follows from [33]; see in particular [33, Table 1].

It is known (see, e. g., [2]) that $\tilde{\mathrm{R}}(T_q) = \tilde{\mathrm{R}}(T_q^{lower}) = q$. Thus we get the following lower bounds on $\omega_u(T_q^{lower}) \geq 2\log(q)/\log(\tilde{\mathrm{S}}(T_q^{lower}))$:

q	Lower Bound on $\omega_u(T_q^{lower})$
2	$2.17795\cdots$
3	$2.16805\cdots$
4	$2.15949\cdots$
5	$2.15237\cdots$

These numbers match the lower bounds obtained by [2, 9] in their study of T_q; our Theorem 3.4 can be viewed as an alternate tool to achieve those lower bounds. The bound approaches 2 as $q \to \infty$, since $\tilde{\mathrm{S}}(T_q^{lower}) = \Omega(q)$ as $q \to \infty$.[8] Interestingly, it is shown in [12, Theorem 4.16] that T_q^{lower} degenerates to T_q over the field $\mathbb{F}_q$, which implies that our bounds above also hold for T_q over $\mathbb{F}_q$.

4.4 The value of the subtensor t_{112} of $CW_q^{\otimes 2}$

A key tensor which arises in applying the Laser Method to increasing powers of CW_q, including [19, 36, 26, 4, 25, 27], is the tensor t_{112} which (for a given positive integer q) is given by

$$t_{112} := \sum_{i=1}^{q} x_{i,0}y_{i,0}z_{0,q+1} + \sum_{k=1}^{q} x_{0,k}y_{0,k}z_{q+1,0} + \sum_{i,k=1}^{q} x_{i,0}y_{0,k}z_{i,k} + \sum_{i,k=1}^{q} x_{0,k}y_{i,0}z_{i,k}.$$

Coppersmith–Winograd [19] and future work studied the value of this tensor. In [19] it is shown that for every $\tau \in [2/3, 1]$,

$$V_\tau(t_{112}) \geq 2^{2/3} q^\tau (q^{3\tau} + 2)^{1/3}.$$

This bound has been used in all the subsequent work using CW_q, without improvement. Here we show it is tight and cannot be improved in the case $\tau = 2/3$:

Proposition 4.1. $V_{2/3}(t_{112}) = 2^{2/3} q^{2/3} (q^2 + 2)^{1/3}$.

Proof. Consider the variable-symmetric tensor $t_s := t_{112} \otimes rot(t_{112}) \otimes rot(rot(t_{112}))$. As in [19], by definition of $V_{2/3}$, for every $\delta > 0$ there is a positive integer n such that $t_s^{\otimes n}$ has a degeneration to $\bigoplus_i \langle a_i, a_i, a_i \rangle$ for values such that $\sum_i a_i^2 \geq (V_{2/3}(T_{112}))^{3n(1-\delta)}$. In particular, by Corollary 2.6 this yields the bound

$$\tilde{\mathrm{S}}(t_s^{\otimes n}) \geq \sum_i a_i^2 \geq (V_{2/3}(t_{112}))^{3n(1-\delta)}.$$

Since this holds for all $\delta > 0$, it follows that $\tilde{\mathrm{S}}(t_s) \geq (V_{2/3}(t_{112}))^3 \geq 2^2 q^2 (q^2 + 2)$.

We now give an upper bound on $\tilde{\mathrm{S}}(t_s)$ using Theorem 3.4. Although we are analyzing t_s, we will make use of a partition of the variables of t_{112}. The partition is as follows: $X_0 = \{x_{i,0} \mid i \in [q]\}$, $X_1 = \{x_{0,k} \mid k \in [q]\}$, $Y_0 = \{y_{i,0} \mid i \in [q]\}$, $Y_1 = \{y_{0,k} \mid k \in [q]\}$, $Z_0 = \{z_{i,k} \mid i,k \in [q]\}$, $Z_1 = \{z_{0,q+1}\}$, and $Z_2 = \{z_{q+1,0}\}$. Hence, $L = \{T_{001}, T_{112}, T_{010}, T_{100}\}$. As in [19], and similar to Proposition 3.8, since t_s is defined as $t_s := t_{112} \otimes rot(t_{112}) \otimes rot(rot(t_{112}))$, it follows that $\tilde{\mathrm{S}}(t_s) \leq \sup_{p \in P(L)} p_X \cdot p_Y \cdot p_Z$. We can

[8] For instance, if we pick the probability distribution p to be the uniform distribution on the $\binom{q}{2}$ terms of T_q^{lower}, then we find $p_X = p_Y = p_Z = \Omega(q)$, and so by Theorem 5.3 below, we get $\tilde{\mathrm{S}}(T_q^{lower}) = \Omega(q)$.

assume, again by symmetry, that any probability distribution p on L assigns the same value v to T_{010} and T_{100}, and the same value $1/2-v$ to T_{001} and T_{112}. We finally get the bound:

$$\tilde{\mathrm{S}}(t_s) \leq \sup_{v\in[0,1/2]} (2q)^2 \cdot \frac{(q^2)^{2v}}{(2v)^{2v}(1/2-v)^{1-2v}}.$$

This is maximized at $v = q^2/(2q^2+2)$, which yields exactly $\tilde{\mathrm{S}}(t_s) \leq 2^2q^2(q^2+2)$. The desired bound follows. □

The only upper bound we are able to prove on V_τ for $\tau > 2/3$ is the straightforward $V_\tau(t_{112}) \leq V_{2/3}(t_{112})^{3\tau/2} = 2^\tau q^\tau (q^2+2)^{\tau/2}$, which is slightly worse than the best known lower bound $V_\tau(t_{112}) \geq 2^{2/3}q^\tau(q^{3\tau}+2)^{1/3}$. It is an interesting open problem to prove tight upper bounds on $V_\tau(T)$ for any nontrivial tensor T and value $\tau > 2/3$. $T = t_{112}$ may be a good candidate since the Laser Method seems unable to improve $V_\tau(t_{112})$ for any τ, even when applied to any small tensor power $t_{112}^{\otimes n}$.

Notice that we were able to prove a tight bound on $\tilde{\mathrm{S}}(t_s)$ here: the upper bound we proved matches a lower bound which we were able to derive from Coppersmith–Winograd's analysis (which made use of the Laser Method) of $V_\tau(t_{112})$. In the next section we will substantially generalize this fact, by showing a tight bound on $\tilde{\mathrm{S}}(T)$ for any tensor T to which the Laser Method applies.

5 Slice rank lower bounds via the Laser Method

In this section, we show that the Laser Method can be used to give matching upper and lower bounds on $\tilde{\mathrm{S}}(T)$ for any tensor T to which it applies. We will build off of Theorem 3.4, which we will show matches the bounds which arise in the Laser Method.

Consider any tensor T which is minimal over X,Y,Z, and let $X = X_1 \cup \cdots \cup X_{k_X}$, $Y = Y_1 \cup \cdots \cup Y_{k_Y}$, $Z = Z_1 \cup \cdots \cup Z_{k_Z}$ be partitions of the three sets of variables. Define T_{ijk}, L, and p_X for a probability distribution p on L, as in the top of Subsection 2.7. Recall in particular that T_{ijk} is T restricted to the sets of variables X_i, Y_j, and Z_k.

Definition 5.1. We say that T, along with partitions of X,Y,Z, is a *laser-ready tensor partition* if the following three conditions are satisfied:

(1) For every $(i,j,k) \in [k_X]\times[k_Y]\times[k_Z]$, either $T_{ijk} = 0$, or else T_{ijk} has a degeneration to a tensor $\langle a,b,c\rangle$ with $ab = |X_i|$, $bc = |Y_j|$, and $ca = |Z_k|$ (i. e., a matrix multiplication tensor which is as big as possible given $|X_i|$, $|Y_j|$, and $|Z_k|$).

(2) There is an integer ℓ such that $T_{ijk} \neq 0$ only if $i+j+k = \ell$.

(3) T is variable-symmetric, and the partitions are T-symmetric.

These conditions are exactly those for which the original Laser Method used by Coppersmith and Winograd [19] applies to T. We note that condition (3) is a simplifying assumption rather than a real condition on T: for any tensor T and partitions satisfying conditions (1) and (2), the tensor $T' := T \otimes rot(T) \otimes rot(rot(T))$ along with the corresponding product partitions, satisfies all three conditions, gives at least as good a bound on ω using the Laser Method as T and the original partitions, and more generally has $\omega_u(T') \leq \omega_u(T)$.

Theorem 5.2 ([19, 21, 36], see also [11, Proposition 15.32])**.** *Suppose T, along with the partitions of X,Y,Z, is a laser-ready tensor partition. Then, for any distribution $p \in P^{sym}(L)$, and any positive integer n, the tensor $T^{\otimes n}$ has a degeneration into*

$$\left(\prod_{i\in[k_X]} p(X_i)^{-p(X_i)}\right)^{n-o(n)} \odot \langle a,a,a\rangle,$$

where

$$a = \left(\prod_{T_{ijk}\in L} |X_i|^{p(T_{ijk})}\right)^{n/2-o(n)}.$$

Proof. Typically, as described in [36, Section 3], there is an additional loss in the size of the degeneration if there are multiple different distributions p, p' with the same marginals (meaning $p(X_i) = p'(X_i)$, $p(Y_j) = p'(Y_j)$, and $p(Z_k) = p'(Z_k)$ for all i, j, k) but different values of $V(p) := \prod_{T_{ijk}\in L} V_\tau(T_{ijk})^{p(T_{ijk})}$ for any $\tau \in [2/3, 1]$. However, because of condition (1) in the definition of a laser-ready tensor partition, the quantity $V(p)$ is equal to

$$\prod_{T_{ijk}\in L} (|X_i|\cdot|Y_j|\cdot|Z_k|)^{p(T_{ijk})\cdot\tau/2},$$

and in particular satisfies $V(p) = V(p')$ for any two distributions p, p' with the same marginals. Thus, we do not incur this loss, and we get the desired degeneration. $\square$

Our key new result about such tensor partitions is as follows:

Theorem 5.3. *Suppose tensor T, along with the partitions of X,Y,Z, is a laser-ready tensor partition. Then,*

$$\tilde{\mathrm{S}}(T) = \sup_{p\in P^{sym}(L)} p_X.$$

Proof. The upper bound, $\tilde{\mathrm{S}}(T) \le \sup_{p\in P^{sym}(L)} p_X$, is given by Proposition 3.8.

For the lower bound, we know from Theorem 5.2 that for all $p \in P^{sym}(L)$, and all positive integers n, the tensor $T^{\otimes n}$ has a degeneration into

$$\left(\prod_{i\in[k_X]} p(X_i)^{-p(X_i)}\right)^{n-o(n)} \odot \langle a,a,a\rangle,$$

where

$$a = \left(\prod_{T_{ijk}\in L} |X_i|^{p(T_{ijk})}\right)^{n/2-o(n)}.$$

By Proposition 2.5, this means $T^{\otimes n}$ has a degeneration to an independent tensor of size

$$\left(\prod_{i\in[k_X]} p(X_i)^{-p(X_i)}\right)^{n-o(n)} \cdot a^2 = p_X^{n-o(n)}.$$

Applying Proposition 2.3 and Proposition 2.4 implies that $\tilde{\mathrm{S}}(T) \ge p_X$ for all $p \in P^{sym}(L)$, as desired. $\square$

Corollary 5.4. *The upper bounds on $\tilde{\mathrm{S}}(CW_{q,\sigma})$, $\tilde{\mathrm{S}}(cw_{q,\sigma})$, $\tilde{\mathrm{S}}(T_q^{lower})$, and $\tilde{\mathrm{S}}(T_q)$ from Section 4 are tight.*

Proof. $CW_{q,\sigma}$, $cw_{q,\sigma}$, and T_q^{lower}, partitioned as they were in the previous section, are laser-ready tensor partitions. The tight bound for T_q follows from the degeneration to T_q^{lower} described in the previous section. □

Corollary 5.5. *Every tensor T with a laser-ready tensor partition (including $CW_{q,\sigma}$, $cw_{q,\sigma}$, and T_q^{lower}) has $\tilde{\mathrm{S}}(T) = \tilde{\mathrm{Q}}(T)$.*

Proof. All tensors satisfy $\tilde{\mathrm{S}}(T) \geq \tilde{\mathrm{Q}}(T)$. In Theorem 5.3, the upper bound on $\tilde{\mathrm{S}}(T)$ showed that $T^{\otimes n}$ has a degeneration to an independent tensor of size $\tilde{\mathrm{S}}(T)^{n-o(n)}$, which implies that $\tilde{\mathrm{Q}}(T) \geq \tilde{\mathrm{S}}(T)$. □

Corollary 5.6. *If T is a tensor with a laser-ready tensor partition, and applying the Laser Method to T with this partition yields an upper bound on ω of $\omega_u(T) \leq c$ for some $c > 2$, then $\omega_u(T) > 2$.*

Proof. When the Laser Method shows, as in Theorem 5.2, that $T^{\otimes n}$ has a degeneration into

$$\left(\prod_{i\in[k_X]} p(X_i)^{-p(X_i)}\right)^{n-o(n)} \odot \langle a,a,a\rangle,$$

the resulting upper bound on $\omega_u(T)$ is that

$$\left(\prod_{i\in[k_X]} p(X_i)^{-p(X_i)}\right)^{n-o(n)} \cdot a^{\omega_u(T)} \geq \tilde{\mathrm{R}}(T)^n.$$

In particular, since the left-hand side equals p_X when $\omega_u(T) = 2$, this yields $\omega_u(T) = 2$ if and only if $p_X = \tilde{\mathrm{R}}(T)$, so if it yields $\omega_u(T) \leq c$, then $\tilde{\mathrm{S}}(T) = p_X < \tilde{\mathrm{R}}(T)^{1-\delta}$ for some $\delta > 0$. Combined with Theorem 2.7 or Theorem 2.9, this means that $\omega_u(T) > 2$. □

A Proof that $\omega_u(CW_{q,\sigma}) \geq 2.16805$ for all q

Define the function $f : [0,1/3] \to \mathbb{R}$ by

$$f(v) := \frac{1}{v^v(2/3-2v)^{2/3-2v}(1/3+v)^{1/3+v}}.$$

In Section 4.1, we showed that

$$\omega_u(CW_{q,\sigma}) \geq \min_{v\in[0,1/3]} 2\frac{\log(q+2)}{\log(q^{2/3-2v}\cdot f(v))}.$$

The value of this optimization problem is computed for $1 \leq q \leq 8$ in a table in Section 4.1, where we see that $\omega_u(CW_{q,\sigma}) \geq 2.16805$ for all $q \leq 8$.

Let v_q denote the argmin for the optimization problem. In particular, for $q = 8$, the argmin is $v_8 = 0.017732422\ldots$. From the $q^{2/3-2v}$ term in the optimization problem, we see that $v_{q+1} \leq v_q$ for all q, and in particular, $v_q \leq v_8$ for all $q > 8$. It follows that $f(v_q) \leq f(v_8) = 2.07389\ldots$ for all $q > 8$. Thus, for all $q > 8$ we have:

$$\omega_u(CW_{q,\sigma}) \geq \min_{v\in[0,1/3]} 2\frac{\log(q+2)}{\log(q^{2/3-2v}\cdot f(v_8))} = 2\frac{\log(q+2)}{\log(q^{2/3}\cdot f(v_8))}.$$

This expression equals $2.18562\ldots$ at $q = 9$, and is easily seen to be increasing with q for $q > 9$, which implies as desired that $\omega_u(CW_{q,\sigma}) \geq 2.16805$ for all $q \geq 9$ and hence all q.

Acknowledgements

I would like to thank Matthias Christandl, Joshua Grochow, Ryan Williams, Virginia Vassilevska Williams, Jeroen Zuiddam, and anonymous reviewers, for many helpful discussions and suggestions.

References

[1] JOSH ALMAN: Limits on the Universal Method for matrix multiplication. In *Proc. 34th Comput. Complexity Conf. (CCC'19)*, pp. 12:1–12:24. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.12] 1

[2] JOSH ALMAN AND VIRGINIA VASSILEVSKA WILLIAMS: Further limitations of the known approaches for matrix multiplication. In *Proc. 9th Innovations in Theoret. Comp. Sci. conf. (ITCS'18)*, pp. 25:1–25:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.ITCS.2018.25] 3, 4, 22, 23

[3] JOSH ALMAN AND VIRGINIA VASSILEVSKA WILLIAMS: Limits on all known (and some unknown) approaches to matrix multiplication. In *Proc. 59th FOCS*, pp. 580–591. IEEE Comp. Soc., 2018. [doi:10.1109/FOCS.2018.00061] 2, 3, 4, 6, 8, 9, 10, 11, 13, 15, 16, 18, 19, 20, 22

[4] JOSH ALMAN AND VIRGINIA VASSILEVSKA WILLIAMS: A refined laser method and faster matrix multiplication. In *Proc. 32nd Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'21)*, pp. 522–539. SIAM, 2021. [doi:10.1137/1.9781611976465.32] 2, 3, 4, 7, 11, 12, 23

[5] NOGA ALON, AMIR SHPILKA, AND CHRISTOPHER UMANS: On sunflowers and matrix multiplication. *Comput. Complexity*, 22(2):219–243, 2013. [doi:10.1007/s00037-013-0060-1] 2

[6] ANDRIS AMBAINIS, YUVAL FILMUS, AND FRANÇOIS LE GALL: Fast matrix multiplication: limitations of the Coppersmith–Winograd method. In *Proc. 47th STOC*, pp. 585–593. ACM Press, 2015. [doi:10.1145/2746539.2746554] 3, 5

[7] DARIO BINI: Border rank of a $p \times q \times 2$ tensor and the optimal approximation of a pair of bilinear forms. In *Proc. 7th Internat. Colloq. on Automata, Languages, and Programming (ICALP'80)*, pp. 98–108. Springer, 1980. [doi:10.1007/3-540-10003-2_63] 11

[8] MARKUS BLÄSER: *Fast Matrix Multiplication*. Number 5 in Graduate Surveys. Theory of Computing Library, 2013. [doi:10.4086/toc.gs.2013.005] 9

[9] 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 Analysis*, 2017:3:1–27. [doi:10.19086/da.1245] 4, 5, 8, 13, 23

[10] JONAH BLASIAK, THOMAS CHURCH, HENRY COHN, JOSHUA A. GROCHOW, AND CHRIS UMANS: Which groups are amenable to proving exponent two for matrix multiplication? 2017. [arXiv:1712.02302] 8

[11] PETER BÜRGISSER, MICHAEL CLAUSEN, AND MOHAMMAD A. SHOKROLLAHI: *Algebraic Complexity Theory*. Volume 315 of *Grundlehren der Math. Wiss.* Springer, 2013. [doi:doi:10.1007/978-3-662-03338-8] 9, 25

[12] MATTHIAS CHRISTANDL, PÉTER VRANA, AND JEROEN ZUIDDAM: Universal points in the asymptotic spectrum of tensors. In *Proc. 50th STOC*, pp. 289–296. ACM Press, 2018. [doi:10.1145/3188745.3188766] 8, 14, 16, 23

[13] MATTHIAS CHRISTANDL, PÉTER VRANA, AND JEROEN ZUIDDAM: Barriers for fast matrix multiplication from irreversibility. In *Proc. 34th Comput. Complexity Conf. (CCC'19)*, volume 137, pp. 26:1–26:17. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.26] 8, 14

[14] MICHAEL B. COHEN, YIN TAT LEE, AND ZHAO SONG: Solving linear programs in the current matrix multiplication time. *J. ACM*, 68(1):3:1–3:39, 2021. Preliminary version in STOC'19. [doi:10.1145/3424305] 4

[15] HENRY COHN, ROBERT KLEINBERG, BALÁZS SZEGEDY, AND CHRISTOPHER UMANS: Group-theoretic algorithms for matrix multiplication. In *Proc. 46th FOCS*, pp. 379–388. IEEE Comp. Soc., 2005. [doi:10.1109/SFCS.2005.39, arXiv:math/0511460] 2

[16] HENRY COHN AND CHRISTOPHER UMANS: A group-theoretic approach to fast matrix multiplication. In *Proc. 44th FOCS*, pp. 438–449. IEEE Comp. Soc., 2003. [doi:10.1109/SFCS.2003.1238217] 2, 11, 22

[17] HENRY COHN AND CHRISTOPHER UMANS: Fast matrix multiplication using coherent configurations. In *Proc. 24th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'13)*, pp. 1074–1086. SIAM, 2013. [doi:10.1137/1.9781611973105.77] 2, 7

[18] DON COPPERSMITH AND SHMUEL WINOGRAD: On the asymptotic complexity of matrix multiplication. *SIAM J. Comput.*, 11(3):472–492, 1982. [doi:10.1137/0211038] 11

[19] DON COPPERSMITH AND SHMUEL WINOGRAD: Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9(3):251–280, 1990. [doi:10.1016/S0747-7171(08)80013-2] 2, 3, 4, 5, 7, 11, 18, 21, 23, 24, 25

[20] ERNIE CROOT, VSEVOLOD F. LEV, AND PÉTER PÁL PACH: Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Ann. Math.*, 185(1):331–337, 2017. [doi:10.4007/annals.2017.185.1.7] 8

[21] ALEXANDER M. DAVIE AND ANDREW JAMES STOTHERS: Improved bound for complexity of matrix multiplication. *Proc. Royal Soc. Edinburgh, Sec. A*, 143(2):351–369, 2013. [doi:10.1017/S0308210511001648] 3, 4, 7, 11, 25

[22] JORDAN S. ELLENBERG AND DION GIJSWIJT: On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. Math.*, 185(1):339–343, 2017. [doi:10.4007/annals.2017.185.1.8] 8

[23] MATTI KARPPA AND PETTERI KASKI: Probabilistic tensors and opportunistic Boolean matrix multiplication. In *Proc. 30th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'19)*, pp. 496–515. SIAM, 2019. [doi:10.1137/1.9781611975482.31] 7

[24] ROBERT KLEINBERG, WILL SAWIN, AND DAVID SPEYER: The growth rate of tri-colored sum-free sets. *Discrete Analysis*, 2018:12:1–10. [doi:10.19086/da.3734] 5

[25] FRANÇOIS LE GALL: Faster algorithms for rectangular matrix multiplication. In *Proc. 53rd FOCS*, pp. 514–523. IEEE Comp. Soc., 2012. [doi:10.1109/FOCS.2012.80] 4, 7, 23

[26] FRANÇOIS LE GALL: Powers of tensors and fast matrix multiplication. In *Proc. 39th Internat. Symp. Symbolic and Algebraic Computation (ISSAC'14)*, pp. 296–303. ACM Press, 2014. [doi:10.1145/2608628.2608664] 2, 3, 4, 7, 11, 12, 23

[27] FRANÇOIS LE GALL AND FLORENT URRUTIA: Improved rectangular matrix multiplication using powers of the Coppersmith–Winograd tensor. In *Proc. 29th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'18)*, pp. 1029–1046. SIAM, 2018. [doi:10.1137/1.9781611975031.67] 4, 7, 8, 23

[28] ERIC NASLUND AND WILL SAWIN: Upper bounds for sunflower-free sets. In *Forum of Mathematics, Sigma*, volume 5, pp. e15:1–10. Cambridge Univ. Press, 2017. [doi:10.1017/fms.2017.12] 8

[29] ARNOLD SCHÖNHAGE: Partial and total matrix multiplication. *SIAM J. Comput.*, 10(3):434–455, 1981. [doi:10.1137/0210032] 11

[30] VOLKER STRASSEN: Gaussian elimination is not optimal. *Numerische Mathematik*, 13(4):354–356, 1969. [doi:10.1007/BF02165411] 2, 10

[31] VOLKER STRASSEN: The asymptotic spectrum of tensors and the exponent of matrix multiplication. In *Proc. 27th FOCS*, pp. 49–54. IEEE Comp. Soc., 1986. [doi:10.1109/SFCS.1986.52] 9, 10

[32] VOLKER STRASSEN: Relative bilinear complexity and matrix multiplication. *J. Reine Angew. Math.*, 1987(375–376):406–443, 1987. [doi:10.1515/crll.1987.375-376.406] 2, 3, 7, 10, 11, 13

[33] VOLKER STRASSEN: Degeneration and complexity of bilinear maps: some asymptotic spectra. *J. Reine Angew. Math.*, 1991(413):127–180, 1991. [doi:10.1515/crll.1991.413.127] 8, 22

[34] TERENCE TAO: A symmetric formulation of the Croot–Lev–Pach–Ellenberg–Gijswijt capset bound, 2016. LINK at terrytao.wordpress.com. 5, 8, 13

[35] TERENCE TAO AND WILL SAWIN: Notes on the "slice rank" of tensors., 2016. LINK at terrytao.wordpress.com. 8, 13, 16

[36] VIRGINIA VASSILEVSKA WILLIAMS: Multiplying matrices faster than Coppersmith–Winograd. In *Proc. 44th STOC*, pp. 887–898. ACM Press, 2012. [doi:10.1145/2213977.2214056] 2, 3, 4, 7, 11, 12, 23, 25

AUTHOR

Josh Alman
Assistant professor
Department of Computer Science
Columbia University
New York, NY, USA
josh@cs.columbia.edu
`http://joshalman.com`

ABOUT THE AUTHOR

JOSH ALMAN is an Assistant Professor of Computer Science at Columbia University. He received his Ph. D. under the co-supervision of Virginia Vassilevska Williams and Ryan Williams at MIT in 2019, then spent two years as a Michael O. Rabin postdoctoral fellow in the Theory of Computation group at Harvard before joining Columbia. In addition to matrix multiplication, Josh is broadly interested in algorithm design and complexity theory, and especially problems in these areas with an algebraic nature. Josh likes solving and creating puzzles, and he is the captain of the ✈✈✈ Galactic Trendsetters ✈✈✈, the team that won the 2020 MIT Mystery Hunt.