A Survey into Lower Bounding Cohn & Umans Matrix Multiplication

Khai Dong

March 15, 2024

Abstract

Bounding matrix multiplication runtime has been an active area of research since 1969 when Strassen showed $\omega \leq 2.81$ [24]. In 2003, Cohn and Umans [13] produced a matrix multiplication framework and conjectured $\omega = 2$. However, a concrete lower bound greater than 2 has never been given. More works over the year by Alman and Williams [4, 3] focus on lower bounding matrix multiplication using various methods including Cohn and Umans'. However, no concrete lowerbound for this method has been given. In this research, we initially aim to give a lower bound tighter than 2 to Cohn & Umans method. Our literature review shows that there are no existing techniques to achieve this task. Therefore, this report summarizes and surveys the current literature regarding this topic.

Acknowledgement

Thanks my advisor and my friends for tolerating me for the duration of this thesis.

Notations

This project uses the following notations:

- $[n] = \{1, 2, \dots, n\}$ where $n \in \mathbb{N}$.
- $[i \dots j] = \{i, i+1, \dots, j\}$ where $i, j \in \mathbb{N}$ and $i \leq j$.
- $\mathbb{N} = \{1, 2, 3, 4, \dots\}$ is the set of positive integers.
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ as the set of integers.
- $\mathbb{Q}$ is the set of rationals.
- $\mathbb{R}$ is the set of reals.
- $\mathbb{C}$ is the set of complex numbers.

Contents

1 Introduction

Matrix multiplication is a fundamental operation in linear algebra that lies at the center of computation in engineering, computer graphics, and quantum physics [16, 1]. As such, over the years researchers have sought ways to improve this operation. Matrix multiplication is relatively short to define (see Section 2.2). However, the question of how fast matrix multiplication can be is a huge challenge. Specifically, in our research, we want to know how fast the Cohn & Umans matrix multiplication can be.

1.1 Prior Works

We define $\mathcal{O}(n^{\omega})$ to be the runtime of the fastest matrix multiplication algorithm for $n \times n$ matrices (see Section 2.2.1 for formal definition). It is widely conjectured that $\omega = 2$.

In 2003, Cohn and Umans [13] proposed a different approach using some results in modern algebra (which are covered in Section 2), conjecturing $\omega = 2$ via their approach. Cohn et al. [14] further built upon their previous work. This framework is the current focus in Professor Anderson's lab, yielding $\omega \leq 2.66$ Anderson et al. [8] and $\omega \leq 2.505$ Anderson and Le [9]. This approach used a group G (see Section 2) to generate a concrete matrix multiplication algorithm with runtime less than $\mathcal{O}(n^3)$. Despite not producing much progress as the current state of the art (the Coppersmith and Winograd identity), the framework has been fully implemented, first partially by Anderson et al. [8], and fully by Dubinsky [19] and Dong [17]. This implementation, however, is inefficient to multiply matrices with usual sizes [17]. Since our greater aim is to obtain $\omega = 2$, this directs our research towards either improving the implementation or obtaining a lowerbound to see if we have hit the limit of this method. This research follows the latter direction, aiming to produce a lower bound greater than 2 for Cohn & Umans' matrix multiplication. A lower bound in this case is the **minimal** upper bound that we can obtain from the Cohn & Umans framework.

1.2 Our Results

However, within prescribed time frame, this research was unable to produce any significant lower bound for Cohn & Umans framework. In the literature, there exists works [2, 22, 5] to produce a lower bound better than 2 for all upper bounds on ω derived from a specific tensor (see Section 2.3). However, these method does not work for the Cohn & Umans framework since it does not use a set tensor to derive upper bounds on ω but uses a structural tensor of an arbitrary group (see Section 3) to derive bounds on ω. In place of an actual lower bound, this paper gives a comprehensive summary on bounding ω (both lower and upper bounding, but focus more into lower bounding). Moreover, this paper have been able to further show the following results.

Proposition. *Let t be a $\mathbb{C}$-tensor over X, Y, Z where $\underline{R}(t) \leq q$ for some integer q. Then, $t \trianglelefteq T_G$ for any abelian group G where $|G| = q$.*

Corollary. *Let t be a $\mathbb{C}$-tensor over X, Y, Z where $\underline{R}(t) \leq q$ for some integer q. Then, $t \trianglelefteq T_G$ for any group G where $|Z(G)| = q$ (the **center** of group G).*

These implies that any of the currently existing argument using the laser method to upper bound ω can be replicated using Cohn & Umans method on an **abelian group** (see Section 2.1.1) G of proper size. This result also implies that arguments using the Universal Method (see Section 2.3.5) on any tensor can also be replicated with proper group G. This result is introduced in Section 5 and is a consequence of the Wedderburn-Artin Theorem (see Section 2.4).

1.3 Paper Outline

The outline of this paper is as follows. Section 2 explains the advanced abstract algebra perquisites. Section 3 describes Cohn & Umans' matrix multiplication framework. The next two sections replicate two results and remark/expand on them:

- No fixed group G may yield an $\mathcal{O}(n^2)$ matrix multiplication algorithm through Cohn & Umans method. This argument has been made by Blasiak et al. [10], Sawin [23], and Blasiak et al. [11], and later generalized by Alman and Vassilevska Williams [3]. These results, however, do not rule out the possibility of an infinite sequence of groups that imply runtimes approaching $\mathcal{O}(n^2)$. The details of this argument are detailed in Section 4.
- Cohn & Umans construction can potentially yield better bound than the current state of the art, the Coppersmith and Winograd tensor. This argument has been made by Alman and Vassilevska Williams [3]. This argument is presented in Section 5. Moreover, in Section 5, we discuss a certain discrepancy in the lower bound results over the years.

Lastly, Section 6 concludes this paper and gives some directions into the upcoming research.

2 Preliminaries

This section served to provide some basics to the reader. For brevity, all proofs are omitted. The same information can also be found in Cormen et al. [16], Bürgisser et al. [12], and Dummit and Foote [20]. We assume the reader knows all basic math including calculus, sets, functions, and equivalent relations. Note that some of the building blocks have been simplified for brevity sake. For precise and rigorous definition, please check with the references above. For ease writing, we omit the citations for the result presented in this section. If the proof is not given, assume it is from the reference above.

2.1 Abstract Algebra Prerequisites

This section provides some of the building blocks for us to understand upcoming subsections and Cohn and Umans' matrix multiplication.

2.1.1 Groups, Rings, and Fields

One example of a group is $\mathbb{Z}$, the set of integers under $+$. We can check that $(\mathbb{Z}, +)$ is indeed a group: the sum of two integers is an integer, the operation $+$ is associative, trivially, the identity elements of $(\mathbb{Z}, +)$ is 0, and for any integer z, $z + (-z) = 0$.

Definition 2.1. *A* ***group*** $(G, *)$ *is a non-empty set of elements defined under a binary operation* $*$ *satisfying the following properties:*

- *closure: for all* $a, b \in G$, $a * b \in G$,
- *associative: for all* $a, b, c \in G$, $(a * b) * c = a * (b * c)$,
- *identity: there exists* $e \in G$ *such that for all* $a \in G$, $a * e = e * a = a$,
- *inverse: for all* $a \in G$, *there exists* $a^{-1} \in G$ *such that* $a * a^{-1} = a^{-1} * a = e$.

When the operation is clear from context, we only write G to stand for the group.

In the context of this project, we only consider finite groups. Informally, a group is an abstraction of familiar mathematical objects. Moreover, we note that $(\mathbb{Z}, +)$ is a rather special case. Note that in this definition, there is no mention of elements commuting through the operation. This means in a group $a * b$ is not necessarily equal to $b * a$. If every element commutes, the group is called **abelian**.

Definition 2.2. *A group G where every element commutes (*$a * b = b * a$ *for all* $a, b \in G$*) is called an* ***abelian*** *group.*

Definition 2.3. *Let G be a group. Define the* ***center*** *of G (denoted* $Z(G)$*) be the set of elements in G that commute with every other element in the group.*

In $(\mathbb{Z}, +)$, only addition and subtraction (adding the inverse) are defined. With integers, we can also multiply. Thus, we define a **ring**, which allows multiplications alongside addition and subtraction.

Definition 2.4. *A **ring** $(R, +, *)$ is a set of elements endowed with 2 binary operations $+$ and $*$ such that*

- *$(R, +)$ is an abelian group,*
- *$*$ is closed and associative,*
- *$*$ is distributive over $+$. That is*

$$\forall a, b, c \in R, (a + b) * c = a * c + b * c \quad \textit{and} \quad c * (a + b) = c * a + c * b.$$

Similar to groups, when the operations are clear from context, we denote R as the ring.

A ring is an abstract object that is closer to our usual number fields. However, some basic properties have yet to be included:

- Whether multiplication is commutative
- Whether the multiplicative identity "1" exists at all
- Whether we can divide (multiply with inverses).

Definition 2.5. *A ring R is **commutative** if $*$ is commutative.*

Definition 2.6. *A ring R is **with unity** if there exists $1 \in R$ such that for all $a \in R$, $1 * a = a * 1 = a$.*

Definition 2.7. *A ring R is a **division ring** if it is with unity and for all non-zero $a \in R$, there exists a^{-1} such that $a * a^{-1} = a^{-1} * a = 1$.*

Definition 2.8. *A **field** is a commutative division ring.*

A field is where we can add, subtract, divide, and multiply, therefore mirroring the basic operations on basic number fields such as $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. We define what it means for 2 rings (or fields) to be the "same".

Definition 2.9. *Two rings R and S are said to be **isomorphic**, denoted $R \cong S$, if there exists an a bijection $\phi : R \to S$ such that for all $a, b \in R$,*

$$\phi(a) +_S \phi(b) = \phi(a +_R b) \quad \textit{and} \quad \phi(a) *_S \phi(b) = \phi(a *_R b).$$

*In this case, ϕ is called a **ring isomorphism**. Since fields are also rings, **field isomorphisms** are ring isomorphisms. If ϕ is not a bijection, but still satisfies the secondary property in Equation 2.9, we call ϕ a **ring homomorphism**.*

Informally, if two rings R and S are isomorphic, we can obtain S by renaming the elements of R using ϕ. This means the two rings are essentially the same object. Here are some examples to build intuition for these objects:

- $\mathbb{Z}$ is a ring with unity but is not a division ring since $2 \in \mathbb{Z}$, but $\frac{1}{2} \notin \mathbb{Z}$.
- $2\mathbb{Z} = \{2z : z \in \mathbb{Z}\}$ is a ring without unity since 1 is odd.
- The set of $n \times n$ matrices is a ring with unity where $*$ is matrix multiplication. Note that matrix multiplication does not commute in general and not all matrices have an inverse.
- The set of $n \times n$ invertible matrices is a division ring, not a field, since matrix multiplication does not necessarily commute.
- $\mathbb{Z}_p$ (integers modulo p) is a field if p is prime. This is a finite field.

2.1.2 Polynomial Rings, Group Algebra, Field Extensions, and Closed Fields

Here, we explore operations that preserve a ring's properties. As in the previous section, these operations mirror what can be done in the usual fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Indeed, the following operation defined how we construct polynomials from a ring.

Definition 2.10. *Let R be a ring and x be an indeterminate. The **polynomial ring** $R[x]$ in x over ring R is the set of all formal sums of the form:*

$$\sum_{i=0}^{n} a_i x^i = a_n x^n + a_{n-1}x^{n-1} + a_{n-1}x^{n-1} + \cdots + a_0$$

*for some $n \geq 0$, $a_i \in R$ for $i \in [0 \ldots n]$, and $a_n \neq 0$. For this polynomial, n is called the **degree** of the polynomial.*

Theorem 2.1. *Let R be a ring and x be an indeterminate. Then, $R[x]$ is a ring.*

Using $R[x]$ as our ring and y as our indeterminate, we can define the polynomial ring $R[x][y]$, also denoted $R[x, y]$, similarly and get another polynomial ring.

Corollary 2.1. *Given a ring R and a finite list of indeterminates x_1, x_2, …, x_n, $R[x_1, x_2, \ldots, x_n]$ is a ring.*

Then, we explore a different property of basic fields like $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. Firstly, we note that $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. We say $\mathbb{R}$ is a **field extension** of $\mathbb{Q}$, and $\mathbb{C}$ is a **field extension** of $\mathbb{R}$.

Definition 2.11. *Let $\mathbb{K}$ and $\mathbb{L}$ be fields. If $\mathbb{L} \subseteq \mathbb{K}$ such that $\mathbb{K}$ and $\mathbb{L}$ share their operations, $\mathbb{K}$ is a **field extension** of $\mathbb{L}$ or, equivalently, $\mathbb{L}$ is a **subfield** of $\mathbb{K}$.*

Consider the polynomial $f(x) = x^2 - 2$ in $\mathbb{Q}[x]$ with roots $\pm\sqrt{2} \notin \mathbb{Q}$. In this case, we say $\pm\sqrt{2}$ are **algebraic** over $\mathbb{Q}$ as $\pm\sqrt{2} \notin \mathbb{Q}$, but are roots of a polynomial in $\mathbb{Q}[x]$. Or more generally, we have the following definition.

Definition 2.12. *Let $\mathbb{K}$ and $\mathbb{L}$ be fields. Suppose $\mathbb{K}$ is a field extension of $\mathbb{L}$. Let $\alpha \in \mathbb{K}$. If α is a root of a polynomial on $\mathbb{L}[x]$, we say α is **algebraic** over $\mathbb{L}$.*

We can then apply the same techniques to fields (which are rings by definition).

Definition 2.13. *Let $\mathbb{L}$ be a field and α be algebraic over $\mathbb{L}$. Then, an **algebraic extension** of $\mathbb{L}$ generated by α, denoted $\mathbb{L}(\alpha)$, is the set of elements of the form*

$$f_0 + f_1\alpha + f_1\alpha^2 + \cdots + f_n\alpha^n$$

where $f_i \in \mathbb{L}$.

Trivially, we can observe that if $\alpha \in \mathbb{L}$, then the expansion would not be interesting as a field is closed under its operations ($\mathbb{L}(\alpha) = \mathbb{L}$). This algebraic extension ”puts” α into $\mathbb{L}$ and allows polynomials with roots α to be factored into $(x - \alpha)$ terms while preserving the characteristic of a field. We can extend $\mathbb{L}$ by all of $\mathbb{L}$'s algebraic elements. Hence, all polynomials $f(x) \in \mathbb{L}(x)$ will factor into polynomials of degree 1, or in another term, **completely split**. This field is called the **algebraic closure** of $\mathbb{L}$, denoted $\overline{\mathbb{L}}$. This mirrors the fact that $\overline{\mathbb{R}} = \mathbb{C}$.

Theorem 2.2. *Let $\mathbb{L}$ be a field. Then $\overline{\mathbb{L}}$ is **algebraically closed**; that is every polynomial $f(x) \in \overline{\mathbb{L}}[x]$ has roots in $\overline{\mathbb{L}}$.*

This also means the only algebraic extension of $\overline{\mathbb{L}}$ is $\overline{\mathbb{L}}$ itself; that is $\overline{\overline{\mathbb{L}}} = \overline{\mathbb{L}}$. An example of an algebraically closed field is $\mathbb{C}$.

Definition 2.14. *Let R be a ring and $G = \{g_1, g_2, \ldots, g_n\}$ be a finite group. $R[G]$ is a **group algebra**: a set of formal sums of the form*

$$a_1g_1 + a_2g_2 + \cdots + a_ng_n \tag{1}$$

*where and $a_i \in R$. Elements of a group algebra **add** and **multiply** like polynomials.*

This construction of a group algebra follows from Corollary 2.1 using elements of a group G as the indeterminates. The difference is that the indeterminates are combined based on group G. Since group G is closed under its operation, all elements of $R[G]$ can be reduced into that in Equation 1.

2.2 Matrices and Matrix Multiplication

This section covers some basic concepts regarding matrix multiplication. Firstly, we define matrices and matrix multiplication.

Definition 2.15. *Let $\mathbb{K}$ be a ring and m, n be positive integers. A $m \times n$* ***matrix*** *is a function $a : [m] \times [n] \to \mathbb{K}$. For shorthand notation, define $a_{i,j} := a(i, j)$.*

Definition 2.16. *Let $\mathbb{K}$ be a ring and n be a positive integer. The set of all matrices $a : [n] \times [n] \to \mathbb{K}$ is defined as the* ***matrix ring*** *$\mathcal{M}_n(\mathbb{K})$.*

We can verify that $\mathcal{M}_n(\mathbb{K})$ is indeed a ring. Informally, we can refer to a matrix as a rectangular array of elements of R arranged into rows and columns. An example of a 2×3 matrix with entries in $\mathbb{Z}$ is

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{or} \quad a : [2] \times [3] \to \mathbb{Z}, \quad a(i, j) = (i - 1) * 3 + j.$$

Definition 2.17. *Let $\mathbb{K}$ be a ring and $a : [m] \times [n] \to \mathbb{K}$ and $b : [n] \times [p] \to \mathbb{K}$ be matrices. Then,* ***matrix multiplication*** *of a and b is the function $a \cdot b : [m] \times [p] \to \mathbb{K}$ where*

$$(a \cdot b)_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

This recovers the same matrix multiplication formula given in linear algebra.

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} \cdot \begin{bmatrix} b_{1,1} & \dots & b_{1,p} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \dots & b_{n,p} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1,k} b_{k,1} & \dots & \sum_{k=1}^{n} a_{1,k} b_{k,p} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{m,k} b_{k,1} & \dots & \sum_{k=1}^{n} a_{m,k} b_{k,p} \end{bmatrix} \tag{2}$$

2.2.1 Exponent ω of Matrix Multiplication

We now discuss the exponent ω that researchers use to gauge the efficiency of a matrix multiplication algorithm.

Definition 2.18. *Let $\mathbb{K}$ be a field. Define*

$$M_{\mathbb{K}}(n) := \textit{the number of multiplication operations over } \mathbb{K} \textit{ to multiply two } n \times n \textit{ matrices.}$$

and

$$\omega(\mathbb{K}) := \inf\{\tau \in \mathbb{R} : M_{\mathbb{K}}(n) = \mathcal{O}(n^{\tau})\}$$

Informally, $\omega(\mathbb{K})$ is the minimum exponent required to multiply two $n \times n$ matrices in field $\mathbb{K}$.

Proposition 2.1. *(Schönhage) If $\mathbb{K}$ is a field extension of $\mathbb{L}$, then $\omega(\mathbb{L}) = \omega(\mathbb{K})$.*

This means $\omega(\mathbb{K})$ is immutable by field extension or restriction. Hence, for this project, we define $\omega := \omega(\mathbb{C})$. Trivially, Equation 2 yields $\omega \leq 3$. It is commonly conjectured that $\omega = 2$.

2.3 Tensors

This section covered some basic information about tensors and how tensors relate to ω.

2.3.1 Basics

Generally, tensors correspond with the set of bilinear maps, one of which is matrix multiplication. The details regarding this correspondence is presented in [12, Chapters 14.1 and 14.2].

Definition 2.19. *Let $\mathbb{K}$ be a field, $X = \{x_1, x_2, \ldots, x_m\}$, $Y = \{y_1, y_2, \ldots, y_n\}$, and $Z = \{z_1, z_2, \ldots, z_p\}$ be formal variables. A $\mathbb{K}$-**tensor** t over X, Y, X is the formal sum*

$$t = \sum_{i,j,k} t_{i,j,k} x_i y_j z_k$$

where $x_i \in X$, $y_i \in Y$, and $z_i \in Z$ and $t_{i,j,k} \in \mathbb{K}$. We denote $t(x_i, y_j, z_k) = t_{i,j,k}$.

Definition 2.20. *Let t be a $\mathbb{K}$-tensor over X, Y, Z. We define the **support** of t as the set of triplets*

$$\mathrm{supp}(t) = \{(x, y, z) : t(x, y, z) \neq 0\} \subseteq X \times Y \times Z.$$

For ease of notation, we assume a field $\mathbb{K}$ for every tensor if not stated otherwise. Then, we define tensor addition and tensor product. These are not rigorous compared to those in [12], but for the sake of this project, this definition is sufficient.

Definition 2.21. *Let $X = \{x_1, x_2, \ldots, x_m\}$, $Y = \{y_1, y_2, \ldots, y_n\}$, and $Z = \{z_1, z_2, \ldots, z_p\}$ be formal variables. Let t and t′ be tensors over X, Y, Z, then*

$$t + t' = \sum_{i,j,k} (t_{i,j,k} + t'_{i,j,k})\, x_i y_j z_k$$

is a tensor over X, Y, Z.

Definition 2.22. *Let $X = \{x_1, x_2, \ldots, x_m\}$, $Y = \{y_1, y_2, \ldots, y_n\}$, $Z = \{z_1, z_2, \ldots, z_p\}$, be formal variables. Define X', Y' and Z' similarly. Let t be a tensor over X, Y, Z and t' be a tensor over X', Y', Z', then*

$$t \times t' = \sum_{i,j,k,i',j',k'} (t_{i,j,k} \cdot t'_{i',j',k'})(x_i y_j z_k \cdot x'_i y'_j z'_k).$$

We assume $(x_i y_j z_k \cdot x'_i y'_j z'_k)$ commutes which means we can collect the x, y and z terms. Then,

$$t \times t' = \sum_{i,j,k,i',j',k'} (t_{i,j,k} \cdot t'_{i',j',k'})(x_i x'_i)(y_j y'_j)(z_k z'_k)$$

is a tensor over $X \times X'$, $Y \times Y'$, $Z \times Z'$.

Two following tensors are particularly useful in this paper.

Definition 2.23. *Let $X = \{x_i\}_{i \in [r]}$, $Y = \{y_i\}_{i \in [r]}$, and $Z = \{z_i\}_{i \in [r]}$ be sets of formal variables. A **vector multiplication tensor** $\langle r \rangle$ is given by*

$$\langle r \rangle := \sum_{i=1}^{r} x_i y_i z_i.$$

Definition 2.24. *Let $X = \{x_{i,j}\}_{(i,j) \in [m] \times [n]}$, $Y = \{y_{j,k}\}_{(j,k) \in [n] \times [p]}$, and $Z = \{z_{i,k}\}_{(i,k) \in [m] \times [p]}$ be sets of formal variables. A **matrix multiplication tensor** $\langle m, n, p \rangle$ is given by*

$$\langle m, n, p \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} x_{i,j} y_{j,k} z_{i,k}.$$

Since we define tensors over formal variables, we can rename the formal variables through a bijection and still get the same tensor. We also have the following proposition.

Proposition 2.2.

$$\prod_{i=1}^{s} \langle m_i, n_i, p_i \rangle = \langle \prod_{i=1}^{s} m_i, \prod_{i=1}^{s} n_i, \prod_{i=1}^{s} p_i \rangle$$

This can be proven by multiplying out all the tensors and collecting the x, y, and z terms in each individual tensor as formal products.

2.3.2 Tensor Rank and Asymptotic Tensor Rank

This sections provides the notion of rank of a tensor, representing the complexity of a tensor (and its corresponding bilinear map).

Definition 2.25. *Let $X = \{x_1, x_2, \dots, x_m\}$, $Y = \{y_1, y_2, \dots, y_n\}$, and $Z = \{z_1, z_2, \dots, y_p\}$, and t be a tensor over X, Y, Z. The **rank** of tensor t, denoted $R(t)$, is the minimum r such that*

$$t = \sum_{s=1}^{r} \Bigg(\sum_{i=1}^{n} \alpha_{s,i} x_i \Bigg) \Bigg(\sum_{j=1}^{m} \beta_{s,j} y_j \Bigg) \Bigg(\sum_{k=1}^{p} \gamma_{s,k} z_k \Bigg)$$

for some $\alpha_i, \beta_j, \gamma_j \in \mathbb{K}$ and $x_i \in X$, $y_j \in Y$, and $z_i \in Z$.

By definition, if we can write t as $\sum_{s=1}^{r} \Bigg(\sum_{i=1}^{n} \alpha_i x_i \Bigg) \Bigg(\sum_{j=1}^{m} \beta_j y_j \Bigg) \Bigg(\sum_{k=1}^{p} \gamma_k z_k \Bigg)$, $R(t) \leq r$.
In 1969, Strassen [24] related the ω to the rank of a matrix multiplication tensor. Strassen's matrix multiplication algorithm is equivalent to the tensor

$$\begin{aligned}\langle 2,2,2\rangle = {} & (x_{11} + x_{22})(y_{11} + y_{22})(z_{11} + z_{22}) + (x_{21} + x_{22})y_{11}(z_{21} - z_{22}) \\ & + x_{11}(y_{12} - y_{22})(z_{11} + z_{22}) + x_{22}(y_{21} - y_{11})(z_{11} + z_{21}) + (x_{11} + x_{12})y_{22}(-z_{11} + z_{12}) \\ & + (x_{21} - x_{11})(y_{11} + y_{12})z_{22} + (x_{12} - x_{22})(y_{21} + y_{22})z_{11}\end{aligned}$$

showing $R(\langle 2,2,2\rangle) \leq 7$, giving $\omega \leq \log_2(7) \approx 2.81$ by Proposition 2.6. Winograd [27] later showed that $R(\langle 2,2,2\rangle) = 7$, giving a tight lowerbound to how fast we can multiply 2×2 matrices.

Proposition 2.3. *Let t_1 and t_2 be tensors over X_1, Y_1, Z_1 and X_2, Y_2, Z_2, respectively. Then,*

$$R(t_1 + t_2) \leq R(t_1) + R(t_2) \quad \textit{and} \quad R(t_1 t_2) \leq R(t_1) R(t_2)$$

This means that the rank is subadditive and submultiplicative. In general, computing the rank is computationally hard for most tensors [21], and we only know the exact rank of a small set of tensors, one of which is the vector multiplication tensor.

Proposition 2.4. *Let $r \in \mathbb{N}$. Then, $R(\langle r\rangle) = r$.*

In this project, our particular interest is in the rank of a matrix multiplication tensor, $R(\langle m, n, p\rangle)$. Proposition 2.5 relates the ranks of matrix multiplication tensors to ω.

Proposition 2.5.

$$\omega := \inf\{\tau \in \mathbb{R} : R(\langle n, n, n\rangle) = \mathcal{O}(n^\tau)\}.$$

The proof for Proposition 2.5 is available in Bürgisser et al. [12]. The following proposition relates ω to the rank of matrix multiplication tensors of non-square matrices.

Proposition 2.6. *If $R(\langle m, n, p\rangle) \leq r$, then $(mnp)^{\omega/3} \leq r$ for positive integers m, n, p, and r.*

However, computing the rank of $\langle m, n, p \rangle$ is computationally hard, thus, making the problem of determining ω through the rank difficult. Therefore, in bounding ω, most research uses an intermediate tensor of known rank. Here we will discuss another type of rank that lets us make use of the definition of ω.

Definition 2.26. *Let t be a tensor over X, Y, Z. The **asymptotic rank** of t, denoted $\tilde{R}(t)$, is defined as*

$$\tilde{R}(t) := \lim_{n \to \infty} R(t^n)^{1/n}.$$

This limit is well-defined. Similar to rank, the asymptotic rank is also subadditive and submultiplicative. In general, $\tilde{R}(t) \leq R(t)$ for any tensor t.

Proposition 2.7.

$$\tilde{R}(\langle m, n, p \rangle) = (mnp)^{\omega/3}$$

A quick intuition is that while we do not know the base of the best matrix multiplication algorithm with runtime $\mathcal{O}(n^\omega)$, the actual matrix we are considering is arbitrarily large by definition of the asymptomatic rank, so we can assume we can multiply this arbitrarily large matrix in $\mathcal{O}(n^\omega)$ time.
Then, we define what it is meant for 2 tensors to be the ”same”.

Definition 2.27. *Let t be a tensor over X, Y, Z and t' be a tensor over X', Y', Z'. Then, t and t' are **isomorphic** (denoted $t \cong t'$) if there exists ring isomorphisms $\alpha : \mathbb{K}[X] \to \mathbb{K}[X']$, $\beta : \mathbb{K}[Y] \to \mathbb{K}[Y']$, and $\gamma : \mathbb{K}[Z] \to \mathbb{K}[Z']$ such that $t(x, y, z) = t'(\alpha(x), \beta(y), \gamma(z))$.*

Proposition 2.8. *Let t be a tensor over X, Y, Z and t' be a tensor over X', Y', Z'. If t and t' are **isomorphic** then*

$$R(t) = R(t') \quad \textit{and} \quad \tilde{R}(t) = \tilde{R}(t').$$

However, deriving the asymptomatic rank is harder than deriving the actual rank of the tensor due to taking the limit of $n \to \infty$. This leads us to the next section.

2.3.3 Tensor Restriction and Combinatorial Restriction

This section explores when a tensor is less complex than another tensor.

Definition 2.28. *Let t and t' be tensors defined over X, Y, Z and X', Y', Z'. t is a **restriction** of t' (denoted $t \leq t'$) if there exists ring homomorphisms $\alpha : \mathbb{K}[X] \to \mathbb{K}[X']$, $\beta : \mathbb{K}[Y] \to \mathbb{K}[Y']$, and $\gamma : \mathbb{K}[Z] \to \mathbb{K}[Z']$ such that $t(x, y, z) = t'(\alpha(x), \beta(y), \gamma(z))$.*

Proposition 2.9. *Let t and t' be tensors defined over X, Y, Z and X', Y', Z'. If $t \leq t'$ and $t' \leq t$, then $t \cong t'$.*

Proposition 2.10. *Let t and t' be tensors defined over X, Y, Z and X', Y', Z'. If $t \leq t'$, then*

$$R(t) \leq R(t') \quad \textit{and} \quad \tilde{R}(t) \leq \tilde{R}(t').$$

Corollary 2.2. *If $\langle m, n, p\rangle \leq t$, $(mnp)^{\omega/3} \leq R(t)$.*

Corollary 2.2 is a natural consequence of Proposition 2.6 and Proposition 2.10. This gives us some way to give bounds to ω through some tensors whose rank is known.

Definition 2.29. *Let t be a tensor defined over X, Y, Z. Let $X' \subseteq X$, $Y' \subseteq Y$, and $Z' \subseteq Z$. Let t' be a tensor defined over X', Y', Z' such that*

$$t'(x, y, z) = \begin{cases} t(x, y, z) & \textit{if } (x, y, z) \in X' \times Y' \times Z' \\ 0 & \textit{otherwise.} \end{cases}$$

Then, t' is a ***combinatorial restriction*** *(or* ***zeroing out****) of t.*

Trivially, if t' is a combinatorial restriction (or zeroing out) of t, $t' \leq t$ via natural projections α, β, and γ.

2.3.4 Tensor Degeneration and Border Rank

Definition 2.30. *Let t, t' respectively be $\mathbb{K}$-tensors over X, Y, Z and X', Y', Z', ϵ be an indeterminate over $\mathbb{K}$, and $q \in \mathbb{N}$. We call t a* ***degeneration of order*** *q of t' (denoted $t \trianglelefteq_q t'$) if there exists a $\mathbb{K}[\epsilon]$-tensor τ over X, Y, Z such that*

$$\epsilon^{q-1}t + \epsilon^q \tau \leq t'.$$

If there exists $q \in \mathbb{N}$ such that $t \trianglelefteq_q t'$, t is said to be a ***degeneration*** *of t' (denoted $t \trianglelefteq t'$).*

Definition 2.31. *Let t be $\mathbb{K}$-tensors over X, Y, Z. Then, we define the* ***border rank*** *of t (denoted $\underline{R}(t)$) is the smallest r such that $t \trianglelefteq \langle r\rangle$.*

By this definition, to show that $\underline{R}(t) \leq r$, it is sufficient to show that

$$\epsilon^{q-1}t + \epsilon^q \tau = \sum_{i=1}^{r} u_i(\epsilon)v_i(\epsilon)w_i(\epsilon)$$

where τ is a $\mathbb{K}[\epsilon]$-tensor, and $u_i(\epsilon), v_i(\epsilon), w_i(\epsilon)$ be elements of in $\mathbb{K}[\epsilon][X]$, $\mathbb{K}[\epsilon][Y]$, and $\mathbb{K}[\epsilon][Z]$, respectively. In particular, for any tensor t, $\tilde{R}(t) \leq \underline{R}(t) \leq R(t)$, and similar to Proposition 2.3, border rank is subadditive and submultiplicative.

Proposition 2.11. *Let t and t' be tensors defined over X, Y, Z and X', Y', Z'. If $t \trianglelefteq t'$, then*

$$\underline{R}(t) \leq \underline{R}(t') \quad \textit{and} \quad \tilde{R}(t) \leq \tilde{R}(t').$$

Proposition 2.10 is a trivial consequence of [12, Lemma 15.24(2)]. Moreover, similar to Proposition 2.6, we can upper bound ω by giving a border rank of a matrix multiplication tensor.

Proposition 2.12. *If $\underline{R}(\langle m, n, p\rangle) \leq r$, then $(mnp)^{\omega/3} \leq r$ for positive integers m, n, p, and r.*

In the studies of ω, we use the border rank of a tensor instead of its rank since by the construction in Definition 2.30, it is possible to construct the border rank of a tensor whose rank can not be determined by our current understandings. In fact, the best known bound on ω is through the use of border rank and degeneration, as we see in Section 5.1.

2.3.5 General Method of Bounding ω using Tensors

Generally, upper bounds on ω are obtained through the Asymptomatic Sum Inequality (Schönhage).

Theorem 2.3. *[12, p. 380] (Asymptomatic Sum Inequality)*

$$\underline{R}\bigg(\sum_{i=1}^{s}\langle m_i, n_i, p_i\rangle\bigg) \leq r \implies \sum_{i=1}^{s}(m_i n_i p_i)^{\omega/3} \leq r.$$

Similarly, we can do the same thing for the asymptotic rank.

Theorem 2.4. *[3, Section 3]*

$$\tilde{R}\bigg(\sum_{i=1}^{s}\langle m_i, n_i, p_i\rangle\bigg) \leq r \implies \sum_{i=1}^{s}(m_i n_i p_i)^{\omega/3} \leq r.$$

To obtain r, we rely on restricting (degenerating) some tensors with known rank (border rank or asymptomatic rank) into a sum of matrix multiplication tensors. Since m_i, n_i, p_i, and r are known, we can solve the inequality for an upperbound on ω. However, for most tensors, deriving for rank is NP-Hard [21]. Thus, in practice, many works [15] rely on border rank and degeneration instead of plain rank and restriction which enables more families of tensors can be used. An example of this is the Coppersmith and Winograd Identity (see Section 5), whose border rank is known. Alman and Vassilevska Williams [3] generalizes these approaches to upper bounding ω with a base tensor t whose rank is known as follows:

- The **Solar Method** (**laser method**) is to zeroing out t or power of t into sum of matrix multiplication tensors $\sum\langle m_i, n_i, p_i\rangle$, yielding

$$\sum\langle m_i, n_i, p_i\rangle \leq t.$$

- The **Galactic Method** is to monomially degenerate (see Section 5) t or power of t into $\sum\langle m_i, n_i, p_i\rangle$, yielding

$$\sum\langle m_i, n_i, p_i\rangle \trianglelefteq t.$$

- Lastly, the **Universal Method** uses a more general degeneration instead a monomial generation.

With known (border/asymptomatic) rank of t, each of the aforementioned restriction/degeneration yields an upper bound on ω. It is worth noting that the Universal method subsumes the Galactic method subsumes the Solar method (laser method).

Remark. It is worth noting that doing the analysis over the asymptomatic rank of t is the same as doing the analysis on the (border) rank of t^n for $n \to \infty$. Most current research relies on taking the n^{th} power of the used tensor t to obtain a larger sum of matrix multiplication tensors on the LHS than analyzing t as is, which gives better upper bound on ω.

Remark. While the Galactic Method and Universal Method are defined, up to the most current work [26], only Solar Method is used.

2.4 Wedderburn-Artin Theorem

Here are the last ingredients to Cohn & Umans matrix multiplication. This report does not discuss what it is meant to be **semisimple**, but the structure of this framework guarantees this property.

Theorem 2.5. *(Wedderburn-Artin) Let $\mathcal{A} = \mathbb{K}[G]$ be a **semisimple** group algebra. Then,*

$$\mathcal{A} \cong \mathcal{M}_{d_1}(D_1) \times \mathcal{M}_{d_2}(D_2) \times \mathcal{M}_{d_3}(D_3) \times \cdots \times \mathcal{M}_{d_k}(D_k)$$

*where d_i's are the **character degrees** of G, and d_i and D_i are uniquely determined by G up to permutations, and D_i's are division rings over $\mathbb{K}$. Addition and multiplication in the RHS are pointwise addition and multiplication in the respective matrix rings.*

This means multiplication in the LHS is the same as pointwise matrix multiplication in the RHS.

Theorem 2.6. *If $\mathbb{K}$ is a field and G is a group, then the group algebra $\mathbb{K}[G]$ is semisimple.*

Since we are working with $\mathbb{C}$ (a field) and a group G in Cohn & Umans method, we are guaranteed that Theorem 2.5 applies, and for all i, $D_i = \mathbb{C}$, since $\mathbb{C}$ is algebraically closed. Thus, we have the following corollary.

Corollary 2.3. *Let G be a group with character degrees $\{d_i\}$. Then,*

$$\mathbb{C}[G] \cong \mathcal{M}_{d_1}(\mathbb{C}) \times \mathcal{M}_{d_2}(\mathbb{C}) \times \mathcal{M}_{d_3}(\mathbb{C}) \times \cdots \times \mathcal{M}_{d_k}(\mathbb{C}). \tag{3}$$

The isomorphism maps the LHS to the RHS is called the ***Wedderburn Decomposition*** Φ.

The following proposition gives us a restriction on what values d_i could be.

Proposition 2.13. *Let G be a group and $\{d_i\}$ be the character degrees of G. Then, $|G| = \sum d_i^2$.*

Proposition 2.13 is true since the dimension must match in Corollary 2.3.

Proposition 2.14. *Let G be an abelian group and $\{d_i\}$ be the character degrees of G. Then, $d_i = 1$ for all i.*

Proposition 2.14 holds since if there exists a character degree greater or equal to 2, then, since square matrix multiplication with size ≥ 2 is known to be non-commutative, there exists non-commuting elements in $\mathbb{C}[G]$, contradicting our condition that G is abelian.

3 Cohn & Umans Matrix Multiplication Framework

This section recounts two approaches to upper bounding ω using a group proposed by Cohn & Umans [13, 14].

Definition 3.1. *Let G be a group and $\mathbb{K}$ be a field. Let $X_G = \{x_g : g \in G\}$, $Y_G = \{y_g : g \in G\}$, $Z_G = \{z_g : g \in G\}$. Then,*

$$T_{\mathbb{K}[G]} := \sum_{g,h \in G} 1_{\mathbb{K}} \cdot x_g y_h z_{gh}$$

is the ***structural tensor*** *of $\mathbb{K}[G]$.*

The structure tensor of $\mathbb{K}[G]$ described the multiplication of elements in the same group algebra given the canonical basis $\{1 \cdot g\}_{g \in G}$.

In both of Cohn & Umans' constructions, the main idea is to reduce the structure tensor of $\mathbb{C}[G]$ into matrix multiplication tensor(s). Since we are only considering $\mathbb{K} = \mathbb{C}$, for ease of notation, we write T_G for the structure tensor of $\mathbb{C}[G]$. Since Cohn & Umans method involves zeroing out the structural tensor T_G, it is a variant of the **laser method**.

Proposition 3.1. *Let G be a group and $\{d_i\}_{i\in[k]}$ be its character degrees. Then,*

$$T_G \cong \sum_{i=1}^{k} \langle d_i, d_i, d_i \rangle.$$

This proof is generally easier using bilinear maps than tensors. We will present the argument using bilinear maps notion of restriction instead of that of tensors. This alternate definition is available in [12, Chapter 14.3] and is equivalent to the one this paper give above.

Definition 3.2. *Let $\phi : U \times V \to W$ and $\phi' : U' \times V' \to W'$ be bilinear maps. A* ***restriction*** *of ϕ' to ϕ is a triple (α, β, γ') of linear maps $\alpha : U \to U'$, $\beta : V \to V'$, and $\gamma' : W' \to W$ such that $\phi = \gamma' \circ \phi' \circ (\alpha \times \beta)$:*

$$\begin{tikzcd}
U \times V \arrow[r, "\phi"] \arrow[d, "\alpha\times\beta"'] & W \\
U' \times V' \arrow[r, "\phi'"] & W' \arrow[u, "\gamma'"']
\end{tikzcd}$$

Denoted, $\phi \leq \phi'$.

Proof. Let $\phi : \mathbb{C}[G] \times \mathbb{C}[G] \to \mathbb{C}[G]$ be the bilinear map represented by T_G. This means ϕ multiplies elements in $\mathbb{C}[G]$. Then, let $\phi' : \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C}) \times \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C}) \to \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$ be the bilinear map to multiply these matrices pairwise. We want to show that $\phi \cong \phi'$ by showing $\phi \leq \phi'$ and $\phi' \leq \phi$.

Let $\Phi : \mathbb{C}[G] \to \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$ be isomorphism described in Corollary 2.3. There exists its inverse Φ^{-1}. We note that both Φ and Φ^{-1} are isomorphisms of their respective domains and co-domains.

Let $(\alpha, \beta, \gamma') = (\Phi, \Phi, \Phi^{-1})$ and $a, b \in \mathbb{C}[G]$ be arbitrary. Then,

$$\gamma' \circ \phi' \circ (\alpha \times \beta)(a, b) = \Phi^{-1} \circ \phi' \circ (\Phi \times \Phi)(a, b) = \Phi^{-1} \circ \phi'(\Phi(a), \Phi(b)).$$

Since Φ is an isomorphism and ϕ' represents multiplication of elements in $\prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$, we have that

$$\phi'(\Phi(a), \Phi(b)) = \Phi(a) \cdot \Phi(b) = \Phi(ab)$$
$$\implies \gamma' \circ \phi' \circ (\alpha \times \beta)(a, b) = \Phi^{-1} \circ \Phi(ab) = ab = \phi(a, b).$$

Since a, b is arbitrary, we have that $\phi \leq \phi'$, by definition.

Then, we take $(\alpha, \beta, \gamma') = (\Phi^{-1}, \Phi^{-1}, \Phi)$ and $a, b \in \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$ be arbitrary. Then,

$$\gamma' \circ \phi \circ (\alpha \times \beta)(a, b) = \Phi \circ \phi \circ (\Phi^{-1} \times \Phi^{-1})(a, b) = \Phi \circ \phi(\Phi^{-1}(a), \Phi^{-1}(b))$$

Since Φ^{-1} is an isomorphism and ϕ represents multiplication of elements in $\mathbb{C}[G]$, we have that

$$\phi(\Phi^{-1}(a), \Phi^{-1}(b)) = \Phi^{-1}(a) \cdot \Phi^{-1}(b) = \Phi^{-1}(ab)$$
$$\implies \gamma' \circ \phi \circ (\alpha \times \beta)(a, b) = \Phi \circ \Phi^{-1}(ab) = ab = \phi'(a, b).$$

Since a, b is arbitrary, we have that $\phi' \leq \phi$, by definition. Thus, we have $\phi \cong \phi'$. This means their corresponding tensors are isomorphic; that is, $T_G \cong \sum_{i=1}^{k} \langle d_i, d_i, d_i \rangle$, as desired. □

Corollary 3.1. *Let G_1 and G_2 be groups whose character degrees are the same. Then, $T_{\mathbb{C}[G_1]} \cong T_{\mathbb{C}[G_2]}$.*

Proof. We admit a global orderings for the character degrees $\{d_i\}$. Let $\phi_1 : \mathbb{C}[G_1] \times \mathbb{C}[G_1] \to \mathbb{C}[G_1]$ and $\phi_2 : \mathbb{C}[G_2] \times \mathbb{C}[G_2] \to \mathbb{C}[G_2]$ be the bilinear maps represented by T_{G_1} and T_{G_2}, respectively, Φ_1 and Φ_2 be the isomorphisms described in Corollary 2.3 for $\mathbb{C}[G_1]$ and $\mathbb{C}[G_2]$, and $\phi' : \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C}) \times \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C}) \to \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$ be the bilinear map to multiply these matrices pairwise. For brevity, we denote $\mathcal{M} := \prod_{i=1}^{k} \mathcal{M}_{d_i}(\mathbb{C})$. Then, we have the diagram:

$$\begin{tikzcd}
\mathbb{C}[G_1] \times \mathbb{C}[G_1] \arrow[r, "\phi_1"] \arrow[d, "\Phi_1 \times \Phi_1"'] & \mathbb{C}[G_1] \\
\mathcal{M} \times \mathcal{M} \arrow[r, "\phi'"] \arrow[d, "\Phi_2 \times \Phi_2'"'] & \mathcal{M} \arrow[u, "\Phi_1^{-1}"'] \\
\mathbb{C}[G_2] \times \mathbb{C}[G_2] \arrow[r, "\phi_2"] & \mathbb{C}[G_2] \arrow[u, "\Phi_2"']
\end{tikzcd}$$

It is trivial to replicate the preceding proof using this diagram. □

3.1 Triple Product Property (TPP) Construction

The TPP construction aims to restrict this tensor to a matrix multiplication tensor.

Definition 3.3. *Let G be a group and $S \subseteq G$. We define the right quotient set $Q(S)$ as*

$$Q(S) := \{s_1 s_2^{-1} : s_1, s_2 \in S\}$$

Definition 3.4. *[13, Definition 2.1] A group G realizes $\langle m, n, p \rangle$ if there exists subsets S, T, U of G such that $|S| = m$, $|T| = n$, and $|U| = p$, and for all $s \in Q(S)$, $t \in Q(T)$, and $u \in Q(U)$,*

$$stu = 1 \implies s = t = u = 1.$$

*We call this condition of S, T, U the **triple product property**. We say G realizes $\langle m, n, p \rangle$ through S, T, U.*

Indeed, the triple product properties enable matrix multiplication through the following embedding into $\mathbb{C}[G]$. Let $a : [m]\times[n] \to \mathbb{C}$ and $b : [n]\times[p] \to \mathbb{C}$ be matrices and admit an ordering of $S-\{s_1, \dots, s_m\}$, $T = \{t_1, \dots, t_n\}$, $U = \{u_1, \dots, u_p\}$, then to compute $a \cdot b$, we embed a and b as

$$\bar{A} = \sum_{(i,j)\in[m]\times[n]} a_{i,j} s_i^{-1} t_j \quad \text{and} \quad \bar{B} = \sum_{(j,k)\in[n]\times[p]} b_{j,k} t_j^{-1} u_k.$$

Then, we can read off the coefficients of matrix $c : [m] \times [p] \to \mathbb{C}$ embed into $\mathbb{C}[G]$ as

$$\bar{C} = \bar{A}\bar{B} = \sum_{(i,k)\in[m]\times[p]} c_{i,k} s_i^{-1} u_k.$$

This means $\langle m, n, p\rangle \leq T_G$; that is we can combinatorially restrict T_G to obtain matrix multiplication tensor $\langle m, n, p\rangle$. To see this, we choose some elements in $\{x_g\}$ to embed A, some elements in $\{y_g\}$ to embed B, choose some elements in $\{z_g\}$ to unembed $C = AB$, and zero out the rest. Moreover, by Proposition 3.1, we have that

$$\langle m, n, p\rangle \leq T_G \cong \sum_{i=1}^{k} \langle d_i, d_i, d_i\rangle$$

Then, by Proposition 2.8, we have that

$$\tilde{R}(\langle m, n, p\rangle) = (mnp)^{\omega/3} \leq \tilde{R}(T_G) = \sum_{i=1}^{k} \tilde{R}(\langle d_i, d_i, d_i\rangle) = \sum_{i=1}^{k} d_i^{\omega}.$$

Cohn & Umans' TPP construction generates an upperbound for ω through the inequality

$$(mnp)^{\omega/3} \leq \sum_{i=1}^{k} d_i^{\omega} \tag{4}$$

for some group G that realizes $\langle m, n, p\rangle$. Note that by producing the associated mapping for Equation 3, we obtain a concrete recursive matrix multiplication algorithm with runtime corresponding to Equation 4; This has been done in Dong [17]. Since $\{d_i\}$ are constants for a specific group G, we obtain a better matrix multiplication algorithm if we find larger S, T, U in G. In fact, if there exists a TPP construction such that $(|S||T||U|)^{2/3} = |G|$, then $\omega = 2$. We can quickly verify this by referring back to Proposition 2.13.

However, finding the subsets S, T, U in a group G that satisfied the triple product property is not an easy task. Therefore, to facilitate findings components for their framework, Cohn et al. [14] proposed a mathematical object called a Strong Unique Solvable Puzzle (SUSP). Informally, a SUSP can be viewed as a $s \times k$ matrix with entries in $\{1, 2, 3\}$ satisfying certain properties. Each SUSP of size s and width q with n_i

entries i (for $i \in \{1, 2, 3\}$) along with a parameter m implies a group G and its subsets S, T, U of size

$$|G| = s! \cdot m^{sq}, \quad |S| = s! \cdot (m-1)^{n_1}, \quad |T| = s! \cdot (m-1)^{n_2}, \quad |U| = s! \cdot (m-1)^{n_3}.$$

This construction also implies that the maximum character degree of G is $s!$. Generally, larger SUSPs imply a better upper bound of ω. Searching and verifying these SUSPs has been a research focus of Anderson et al. [8] and Anderson and Le [9].

3.2 Simultaneous Triple Product Property (STPP) Construction

Following a similar idea to the TPP construction, the STPP construction aims to realize multiple matrix multiplication tensors through a single group G. Similar to Equation 4, by Theorem 2.4, we obtain ω through the inequality

$$\sum (m_i n_i p_i)^{\omega/3} \leq \sum_{i=1}^{k} d_i^{\omega} \tag{5}$$

where

$$\sum \langle m_i, n_i, p_i \rangle \leq T_G \cong \sum_{i=1}^{k} \langle d_i, d_i, d_i \rangle.$$

This is another attempt to obtain $\omega = 2$ since this results in using more parts of the group G. Rather than using 3 subsets of G, the STTP construction uses multiple triplets of subsets of G.

Definition 3.5. *[14] We says ℓ triplet of subsets S_i, T_i, U_i where $|S_i| = m_i$, $|T_i| = n_i$, $|U_i| = p_i$ and $i \in [\ell]$ of a group G satisfy the simultaneous triple product property if*

- *S_i, T_i, U_i satisfy the triple product property for every $i \in [\ell]$.*
- *For every $(i, j, k) \in [\ell]^3$,*

$$s_i(s_j)^{-1} t_j (t_k)^{-1} u_k (u_i)^{-1} = 1 \implies i = j = k,$$

where $s_i \in S_i$, $s_j \in S_j$, $t_j \in T_j$, $t_k \in T_k$, $u_k \in U_k$, and $u_i \in U_i$.

Then, we say the group G simultaneously realizes $\langle m_1, n_1, p_1 \rangle, \ldots, \langle m_\ell, n_\ell, p_\ell \rangle$.

This construction gives us a bound on ω by Equation 5. Since this construction realizes several matrix multiplications simultaneously, this method is only practical to obtain theoretical bounds on ω. We observe this construction subsumes the TPP construction and potentially gives better bounds on ω.

4 Lower Bound on Cohn & Umans Method

This section is to present an overview of the argument where it is not possible to achieve $\omega = 2$ with any fixed group G using Cohn & Umans method. This results are from [10, 23, 11] and are later generalized by [3].

4.1 Tri-colored Sum-Free Set

The result in this section relies on an upper bound on the size of a tri-colored sum-free set in group G.

Definition 4.1. *Let G be a group. A **tricolored sum-free set** in G is a set $M \subseteq G^3$ such that*

- *for all $(a, b, c) \in M$, $ab = c$, and*
- *for all $(a_1, b_1, c_1), (a_2, b_2, c_2), (a_3, b_3, c_3) \in M$ which are not the same triple, we have $a_1 b_1 \neq c_3$.*

For any group G, the maximal size of the tri-colored sum-free set determines the minimal bound on ω can be obtained; the details are presented in the next section. There are also equivalent definition that we are going to use interchangeably.

Definition 4.2. *Let S, T, U be sets. A **perfect matching** M of S, T, U is a subset of $S \times T \times U$ such that the projection maps $(s, t, u) \mapsto s$, $(s, t, u) \mapsto t$, $(s, t, u) \mapsto u$ are bijective.*

Definition 4.3. *Let $(G, +)$ be a group. A **tricolored sum-free set** in $(G, +)$ is a triple of subsets $S, T, U \subseteq G$ such that*

$$M = \{(s, t, u) \in S \times T \times U : s + t + u = 0\} \tag{6}$$

*is a perfect matching in S, T, U and 0 is the identity of G. The **cardinality** of the tricolored sum-free set is $|M| = |S| = |T| = |U|$.*

Theorem 4.1. *(Sawin [23, Theorem 1]) Let G be a non-trivial finite group. There exists a constant $\delta_G < 1$ such that for any positive integer n, any tri-colored sum-free set in G^n has size at most $(\delta_G |G|)^n$.*

Sawin [23] work expanded on Blasiak et al. [10], holding for all non-trivial group instead of just the abelian group. Blasiak et al. [11] would later shows that certain families on groups can not yield $\omega = 2$ using similar construction.

4.2 Independent Tensor

This section defines independent tensor and link it to the defintion of tri-colored sum-free set.

Definition 4.4. *Let $(x, y, z), (x', y', z') \in X \times Y \times Z$. We say they are* ***independent*** *if $x \neq x'$, $y \neq y'$, and $z \neq z'$.*

Definition 4.5. *Let t be a $\mathbb{K}$-tensor over X, Y, Z. We say t is* ***independent*** *if $t(x_i, y_j, z_k) \neq 0$ and $t(x_{i'}, y_{j'}, z_{k'}) \neq 0$ and $(i, j, k) \neq (i', j', k')$, then (i, j, k) and (i', j', k') are independent.*

We note that $\langle q \rangle$ is independent. Indeed, an independent tensor with support of size q is isomorphic to $\langle q \rangle$. Bringing this back to the notion of border rank, we can observe that if there exists a tricolored sum set of size q in G, then $\langle q \rangle$ is a zeroing out of T_G; this is true since Definition 4.1 and 4.3 are equivalent [10, 23].

Definition 4.6. *Let t be a $\mathbb{K}$-tensor over X, Y, Z. We define the* ***independent number*** *of t, denoted $I(t)$, to be the maximal size of the independent tensor that can be degenerated from t.*

Similar to rank, we define the asymptotic independent number, $\tilde{I}(t)$.

Definition 4.7. *Let t be a $\mathbb{K}$-tensor over X, Y, Z. We define the* ***asymptotic independent number*** *of t, denoted $\tilde{I}(t)$, to be*

$$\tilde{I}(t) = \lim_{n \to \infty} I(t^n)^{1/n}.$$

Remark. If $\tilde{I}(t) = \tilde{R}(t)$, then $\omega = 2$, and if $\tilde{I}(t) < \tilde{R}(t)$, then analysing such tensor t can not yield $\omega = 2$ using the Galactic method [3] (See Section 2.3.5). For ease of notation, we define $\omega_g(t)$ to be the minimal upper bound that can be obtained from applying the Galactic method on t.

Theorem 4.2. *[3, Theorem 6.1] For any finite group G, $\omega_g(T_G) > 2$.*

Proof. Let $\delta < 1$ be the constant in Theorem 4.1. We have that G^n has tri-colored sum free set of size at most $(\delta|G|)^n$. Therefore,

$$\tilde{I}(T_G)) = \lim_{n \to \infty} I(T_G^n)^{1/n} = \lim_{n \to \infty} I(T_{G^n})^{1/n} \leq \lim_{n \to \infty} \left((\delta|G|)^n\right)^{1/n} = \delta|G| < |G| \leq \sum d_i^\omega = \tilde{R}(T_G).$$

This means $\tilde{I}(T_G) < \tilde{R}(T_G)$, implying $\omega_g(T_G) > 2$. □

Since the Galactic method subsumes the solar method (subsumes Cohn & Umans), we can not get $\omega = 2$ from a fixed group G.

Remark. Note that this result does not imply the impossibility of obtaining $\omega = 2$ using Cohn & Umans method. It is still possible to obtain $\omega = 2$ if there exists a family of groups G_α for $\alpha \in I$ where $\sup\{\tilde{I}(G_\alpha)\} = \inf\{\tilde{R}(G_\alpha)\}$, which means we can multiply matrices infinitely close to $\mathcal{O}(n^2)$.

5 Analyzing Group Tensor Subsumes Analysing $\mathbb{T}_q$

This section shows how we can lower bound the upper bounds yielded by Cohn&Umans' method.

5.1 The Coppersmith-Winograd (CW) Identity

This section introduces the Coppersmith and Winograd Identity, and the method to obtain better bound on ω from this particular tensor.

Definition 5.1. *Let $q \geq 1$ be an integer. Then, we defined the Coppersmith-Winograd Identity parameterized by q as the tensor*

$$\mathbb{T}_q = x_0 y_0 z_{q+1} + x_{q+1} y_0 z_0 + x_0 y_{q+1} z_0 + \sum_{i=1}^{q} \left(x_i y_0 z_i + x_0 y_i z_i + x_i y_i z_0 \right)$$

over $X = \{x_i\}_{i \in [0 \ldots q+1]}$, $Y = \{y_i\}_{i \in [0 \ldots q+1]}$, and $Z = \{z_i\}_{i \in [0 \ldots q+1]}$.

Analyzing this tensor has yielded the best upper bounds on ω so far [15, 18, 26]. This section provides the elementary argument provided by Coppersmith and Winograd [15] and explains what general improvements can be made to obtain better bounds on ω.

Let $q \geq 1$ be an integer. We consider the tensor $\mathbb{T}_q$ over the X, Y, Z defined in Definition 5.1 and the decomposition D, where we partition X, Y, Z into

$$\begin{aligned} X &= X_0 \cup X_1 \cup X_2 = \{x_0\} \cup \{x_1, x_2, \ldots, x_q\} \cup \{x_{q+1}\}, \\ Y &= Y_0 \cup Y_1 \cup Y_2 = \{y_0\} \cup \{y_1, y_2, \ldots, y_q\} \cup \{y_{q+1}\}, \\ Z &= Z_0 \cup Z_1 \cup Z_2 = \{z_0\} \cup \{z_1, z_2, \ldots, z_q\} \cup \{z_{q+1}\}. \end{aligned}$$

Then, we observe that

$$\mathbb{T}_q = x_0^{[0]} y_0^{[0]} z_{q+1}^{[2]} + x_{q+1}^{[2]} y_0^{[0]} z_0^{[0]} + x_0^{[0]} y_{q+1}^{[2]} z_0^{[0]} + \sum_{i=1}^{q} \left(x_i^{[1]} y_0^{[0]} z_i^{[1]} + x_0^{[0]} y_i^{[1]} z_i^{[1]} + x_i^{[1]} y_i^{[1]} z_0^{[0]} \right)$$

where we use $x_i^{[j]}$ to denote $x_i \in X_j$. We denote $y_i^{[j]}$ and $z_i^{[j]}$ similarly. We can rewrite:

$$\begin{aligned} \mathbb{T}_q &= x_0^{[0]} y_0^{[0]} z_{q+1}^{[2]} + x_{q+1}^{[2]} y_0^{[0]} z_0^{[0]} + x_0^{[0]} y_{q+1}^{[2]} z_0^{[0]} + \sum_{i=1}^{q} x_i^{[1]} y_0^{[0]} z_i^{[1]} + \sum_{i=1}^{q} x_0^{[0]} y_i^{[1]} z_i^{[1]} + \sum_{i=1}^{q} x_i^{[1]} y_i^{[1]} z_0^{[0]} \\ &\cong \langle 1,1,1 \rangle^{[0,0,2]} + \langle 1,1,1 \rangle^{[2,0,0]} + \langle 1,1,1 \rangle^{[0,2,0]} + \langle q,1,1 \rangle^{[1,0,1]} + \langle 1,1,q \rangle^{[0,1,1]} + \langle 1,q,1 \rangle^{[1,1,0]} \end{aligned}$$

where $\langle m,n,p \rangle^{[i,j,k]}$ denote the matrix multiplication tensor $\langle m,n,p \rangle$ over X_i, Y_j, Z_k. Moreover, $\mathbb{T}_q \trianglelefteq \langle q+2 \rangle$, implying $\underline{R}(\mathbb{T}_q) \leq q+2$ by

$$\sum_{i=1}^{q} \epsilon^{-2}\big(x_0+\epsilon x_i\big)\big(y_0+\epsilon y_i\big)(z_0+\epsilon z_i) - \epsilon^{-3}\left(x_0+\epsilon^2\sum_{i=1}^{q} x_i\right)\left(y_0+\epsilon^2\sum_{i=1}^{q} y_i\right)\left(z_0+\epsilon^2\sum_{i=1}^{q} z_i\right)$$

$$(\epsilon^{-3}-q\epsilon^{-2})(x_0+\epsilon^3 x_{q+1})(y_0+\epsilon^3 y_{q+1})(z_0+\epsilon^3 z_{q+1}) = \mathbb{T}_q+\epsilon G,$$

where G is a $\mathbb{K}[\epsilon]$-tensor over X, Y, Z. In fact, $\underline{R}(\mathbb{T}_q) = q+2$. Coppersmith and Winograd [15] showed how to **zero out** $(\mathbb{T}_q)^n$ into sum of matrix multiplication tensors (**laser method**). Through this construction, Theorem 2.4 implies $\omega \leq 2.41$. Coppersmith further shows that analyzing $(\mathbb{T}_q^2)^n$ yields slightly better bounds. A detailed argument is available in [15, 7]. It was expected that analyzing a larger power of $\mathbb{T}_q$ would potentially yield $\omega \leq 2$. However, Ambainis et al. [6] shows that the best upper bound on ω that can be obtained by analyzing any power of $\mathbb{T}_5$ is 2.3078.

Remark. During my survey, discrepancies were coming up in the literature regarding this limit. This result first showed up in Ambainis and Filmus [5] and Ambainis et al. [6]. In May 2014, Ambainis and Filmus [5] claimed that the best bound on ω obtained by analyzing the n^{th} power of $\mathbb{T}_5$ using the laser method could be is $\omega \leq 2.3078$ and explicitly stated that the such lower bound for $q=2$ is $\omega \leq 2.254$. In Ambainis et al. [6], the claim becomes

> *taking the Nth power tensor of the Coppersmith-Winograd Identity cannot yield an algorithm with running time $\mathcal{O}(n^{2.3078})$ for any value of N*

omitting $q=5$. In Vassilevska Williams et al. [26], citing [6], the claim generalizes further to

> *All work on matrix multiplication since 1986 [...] has used various variants of the so-called laser method. The strongest limitation result known for the laser method and its variants [[6]] is such techniques can not yield $\omega < 2.3078$.*

As far as this paper has been surveying the current literature, such bound is only specific to the power of $\mathbb{T}_5$ and does not apply to any other variants of the laser method. The author does not see the logical basis for how this result has been generated over time and hopes to devote more time to clarify the situation since this has caused a potential misinterpretation of the lower bound result presented in the literature [4, 3]. While it is true that $\mathbb{T}_5$ gives the best analysis so far (by an optimization from Coppersmith and Winograd [15]), it is currently unknown to the author why the lower bound for $\mathbb{T}_5$ may apply for Cohn & Umans method (or any variants of the laser method) as no justification has been given. Therefore, the author chooses to incorporate such statements into its content.

On the other hand, Alman [2] has worked out the limit of analyzing $\mathbb{T}_q$ using the Universal Method, being $\omega \leq 2.16805$ be the minimal bound that could be obtained. This method also works on lower bounding all upperbounds obtained by analyzing a set tensor using the Universal Method.

Moving along with the subsumption of Coppersmith and Winograd by group tensors, we introduced the following tensor to show that analyzing a group tensor can potentially yield at least as good as analyzing $\mathbb{T}_q$.

Definition 5.2. *Let $q \geq 1$ be an integer. Then, we defined the rotated Coppersmith-Winograd Identity parameterized by q as the tensor*

$$\mathbb{T}_q^\sigma = x_0 y_0 z_{q+1} + x_{q+1} y_0 z_0 + x_0 y_{q+1} z_0 + \sum_{i=1}^{q} \left(x_i y_0 z_i + x_0 y_i z_i + x_i y_{\sigma(i)} z_0 \right)$$

over $X = \{x_0, \ldots, x_{q+2}\}$, $Y = \{y_0, \ldots, y_{q+2}\}$, and $Z = \{z_0, \ldots, z_{q+2}\}$ where σ is a permutation of $\{1, \ldots, q\}$.

Alman and Vassilevska Williams [3] showed that analyzing $\mathbb{T}_q^\sigma$ yields the exact same bound on ω as $\mathbb{T}_q$. In this paper, we introduce $\mathbb{T}_q^\sigma$ since $\mathbb{T}_q$ can not be monomially degenerate from an arbitrary T_G.

5.2 Monomial Degeneration

This section shows a special type of degeneration called **monomial degeneration** (also called **combinatorial degeneration**).

Definition 5.3. *Let t, t' be $\mathbb{K}$-tensors over X, Y, Z. We say t is a monomial degeneration of t' if $t \subseteq t'$ and there exists function $a : X \to \mathbb{Z}$, $b : Y \to \mathbb{Z}$, and $c : Z \to \mathbb{Z}$ such that*

- $t'(i,j,k) \neq 0 \implies a(x_i) + b(y_j) + c(z_k) \geq 0$,
- $a(x_i) + b(y_j) + c(z_k) = 0 \implies t(i,j,k) = t'(i,j,k)$,
- $t(i,j,k) \neq 0 \implies a(x_i) + b(y_j) + c(z_k) = 0$.

It can be proven that the existence of such a, b, c implies $t \trianglelefteq t'$ [12, Chapter 15]. Moreover, a zeroing out is monomial degeneration with $a(x) \geq 0$ for all $x \in X$, $b(y) \geq 0$ for all $y \in Y$, and $c(z) \geq 0$ for all $z \in Z$. We can think of this as zeroing out any variable x, y, z with $a(x) \geq 0$ or $b(y) \geq 0$ or $c(z) \geq 0$. This gives us an easier way to describe a monomial degeneration without relying on the definition.

5.3 Monomially Degenerate T_G into $\mathbb{T}_{|G|-2}$ and More

Theorem 5.1. *[3, Theorem 7.2] For every finite group G of order $|G| = q$, there is a monomial degeneration of T_G into a tensor $t \cong \mathbb{T}_{|G|-2}^\sigma$.*

Proof. Let G be a group of order q. Let $e \in G$ be the identity and $g \in G$ be any other element. We define the maps $a : X_G \to \mathbb{Z}$, $b : Y_G \to \mathbb{Z}$, and $z : Z_G \to \mathbb{Z}$ as follows:

- $a(x_e) = b(y_e) = c(z_e) = 0$,
- $a(x_g) = b(y_g) = -c(z_g) = 2$,
- $a(x_h) = b(y_h) = -c(z_h) = 1$ for all $h \in G \setminus \{e, g\}$.

Let t be the monomial degeneration of T_G defined by a, b, c. We define $\sigma : G \setminus \{e, g\} \to G \setminus \{e, g\}$ where $\sigma(h) = h^{-1}g$; σ is a permutation of $G \setminus \{1, g\}$ since $\sigma(h_1) = \sigma(h_2) \implies h_1^{-1}g = h_2^{-1}g \implies h_1^{-1} = h_2^{-1} \implies h_1 = h_2$ and for all $h_0 \in G \setminus \{e, g\}$, $\sigma(gh_0^{-1}) = (gh_0^{-1})^{-1}g = h_0 g^{-1} g = h_0$ where $gh_0^{-1} \in G \setminus \{e, g\}$ since $h_0 \in G \setminus \{e, g\}$. Then, we can see that

- $x_e y_e z_e \in t$ since $a(x_e) = b(y_e) = c(z_e) = 0$.
- $x_e y_h z_h \in t$ for all $h \in G \setminus \{e\}$ since $a(x_e) + b(y_h) + c(y_h) = 0 + 1 - 1 = 0$.
- $x_h y_e z_h \in t$ for all $h \in G \setminus \{e\}$ since $a(x_h) + b(y_e) + c(y_h) = 1 + 0 - 1 = 0$.
- $x_h y_{\sigma(h)} z_g \in t$ for all $h \in G \setminus \{e, g\}$ since $a(x_h) + b(\sigma(h)) + c(z_g) = 1 + 1 - 2 = 0$.

Moreover,

- $x_{h_1} y_{h_2} z_{h_3} \notin t$ for $h_1, h_2, h_3 \in G \setminus \{e, g\}$ with $h_3 = h_1 h_2$ since $a(h_1) = b(h_2) = 1$ and $c(h_3) = -1$.
- $x_h y_{h^{-1}} z_e \notin t$ for $h \in G \setminus \{e, g\}$ since $a(x_h) = b(y_{h-1}) = 1$ and $c(z_e) = 0$.
- $x_g y_{h_1} z_{h_2} \notin t$ for $h \in G \setminus \{e, g\}$ with $gh_1 = h_2$ since $a(x_g) = 2$, $b(y_{h_1}) = 1$ and $c(z_{h_2}) = -1$.
- $x_{h_1} y_g z_{h_2} \notin t$ for $h \in G \setminus \{e, g\}$ with $gh_1 = h_2$ since $a(x_{h_1}) = 1$, $b(y_g) = 2$ and $c(z_{h_2}) = -1$.
- $x_g y_{g^{-1}} z_e \notin t$ since $a(g) = 2$, $b(g^{-1}) = 1$ or 2 and $c(z_e) = 0$, so the sum is at least 3.
- $x_{g^{-1}} y_g z_e \notin t$ since $a(g) = 1$ or 2, $b(g^{-1}) = 2$ and $c(z_e) = 0$, so the sum is at least 3.
- $x_g y_g z_{g^2} \notin t$ since $a(x_g) = b(y_g) = 2$ and $c(z_{g^2}) \geq -2$, so the sum is at least 2.

This covers all entries of T_G, yields

$$t = x_e y_e z_e + x_e y_g z_g + x_g y_e z_g + \sum_{h \in G \setminus \{e, g\}} \left(x_e y_h z_h + x_h y_e z_h + x_h y_{\sigma(h)} z_e \right) \cong \mathbb{T}^{\sigma}_{q-2},$$

since $|G \setminus \{e, g\}| = q - 2$, as desired. □

> (Alman and Vassilevska Williams [3]) *Remark 7.1. An immediate consequence of this monomial degeneration is that applying any implementation of the Solar, Galactic or Universal method on T_G for any finite group G with $\tilde{R}(T_G) = |G|$ yields the same upper bounds on ω as the best known analysis of* $\mathbb{T}_{|G|-2}$.

By Theorem 5.1, we have that

$$\mathbb{T}^{\sigma}_{|G|-2} \trianglelefteq T_G$$

for any group G and suitable σ. Thus, we can see that if $\tilde{R}(T_G) = |G|$, analyzing T_G using Cohn & Umans' method would yield the same ω than analyzing $\mathbb{T}^{\sigma}_{|G|-2}$ [3, Remark 7.1]. However, we acknowledge that analyzing T_G could potentially yield a better upper bound on ω if T_G can be zeroed out/degenerated into a larger sum of matrix multiplication tensors.

Remark. This result is a generalized of a result from Alman and Williams [4], where $G = \mathbb{Z}_q$, the group of integer mod q. Moreover, Remark 7.1 [3] may be misinterpreted as analyzing T_G can only yield the same bound as $\mathbb{T}^{\sigma}_{|G|-2}$. The author admitted his inadequacy while reading this paper, and previously thought that this means we can not get $\omega < 2.3078$ using Cohn&Umans method. However, it is the other way around where analyzing T_G would subsume analyzing $\mathbb{T}_q$. The author further acknowledges that such results are mentioned in the abstract and introduction of Alman and Williams [4]:

> *We also prove a number of complementary results along the way, including that for any group G, the structural tensor of $\mathbb{C}[G]$ can be used to recover the best bound on ω which the Coppersmith-Winograd approach gets using $\mathbb{T}_{|G|-2}$ as long as the asymptotic rank of the structural tensor is not too large.*

> *Every finite group G has a monomial degeneration to some generalized Coppersmith and Winograd tensor of parameter $q = |G| - 2$. Thus, applying the Galactic method on T_G for every G (with sufficiently small asymptotic rank, i.e. $\tilde{R}(T_G) = |G|$) can yield the current best bounds on ω.*

This paper initially interprets these quotes as Cohn & Umans can never perform better than analyses on Coppersmith and Winograd identity. However, this quote only shows the possibility of obtaining the same ω by zeroing out a group tensor and its powers and **does not** indicate that Cohn & Umans approach is worse than Coppersmith and Winograd's.

Linking the result back to Cohn & Umans method since the current best research on $\mathbb{T}_q$ relies on zeroing out $\mathbb{T}_q^n$ into a sum of matrix multiplication tensor, this paper suspects the possibility of obtaining the same current best bound of 2.371552 [26] or better using Cohn & Umans method. Moreover, we prove the following result, deriving from [3], where a great number of tensors used in matrix multiplication research can be monomially degenerated from T_G where G is abelian, which will give the corresponding (border) rank.

Proposition 5.1. *Let t be a $\mathbb{C}$-tensor over X, Y, Z where $\underline{R}(t) \leq q$ for some integer q. Then, $t \trianglelefteq T_G$ for any abelian group G where $|G| = q$.*

Proof. By definition, $\underline{R}(t) \leq q$ implies $t \trianglelefteq \langle q \rangle$. It is sufficient to show that $\langle q \rangle \trianglelefteq T_G$.
Let $\{d_i\}_{i \in [k]}$ be the character degrees of G. By Proposition 3.1, $T_G \cong \langle d_i, d_i, d_i \rangle$. Moreover, since G is abelian, by Proposition 2.14, $d_i = 1$ for all i and $k = |G|$. Thus, the structural tensors of T_G is isomorphic to the structural tensor of $\sum_{i=1}^{|G|} \langle 1, 1, 1 \rangle = \sum_{i=1}^{|G|} \langle 1 \rangle$; that is

$$T_G \cong \sum_{i=1}^{|G|} \langle 1, 1, 1 \rangle \cong \sum_{i=1}^{|G|} \langle 1 \rangle$$

where the variables X, Y, Z of each of $\langle 1 \rangle$ is disjoint pairwise. Since $T_G \cong \langle |G| \rangle$, $\langle q \rangle \trianglelefteq T_G$, as desired. □

Corollary 5.1 shows a slightly more general result than Proposition 5.1.

Corollary 5.1. *Let t be a $\mathbb{C}$-tensor over X, Y, Z where $\underline{R}(t) \leq q$ for some integer q. Then, $t \trianglelefteq T_G$ for any group G where $|Z(G)| = q$ (the* ***center*** *of group G).*

Proof. Since $\underline{R}(t) \leq q$ and $Z(G)$ is a group, $t \trianglelefteq T_{Z(G)}$, by Proposition 5.1. Trivially, $T_{Z(G)}$ is a zero out of T_G where we zero out all non-commuting elements. Thus, $t \trianglelefteq T_{Z(G)} \trianglelefteq T_G$, as desired. □

This shows that any argument ever made using Universal Method (including laser method) can be replicated using a structural tensor of an abelian group. This is on the same line of [4, 3] where they show some known tensors used in [15, 25] can be degenerated from T_G. This result is more general as it would work for any known tensor t, but less general in the way that is only for abelian group. Theoretically, this results show that if we can formulate a ”triple product property” which made use of degeneration, it is possible to obtain the current best bound on ω with sufficiently small group G.

6 Conclusion

This research has failed its initial directive to give any lower bound tighter than 2 for Cohn & Umans method. Therefore, this report aims to summarize the current techniques to lower bound matrix multiplication techniques, give a quick overview of the techniques used to lower bound upper bounds on ω using specific methods, and present a reference list to be used in future research.

Regarding the lower bounding Cohn & Umans method, this research has discovered no existing method in the literature that can produce a lower bound better than 2 for Cohn & Umans method. One way to lower bound this method could be by coming up with a better construction for a large tri-colored sum-free set in G^n. However, as pointed out in Section 4, this construction of a tri-colored sum-free set in G^n is likely tight.

At the same time, the other lines of this thesis can continue to pursue generating a better matrix multiplication algorithm to obtain better bounds on ω that we previously obtained and/or further improve Dubinsky [19] and Dong [17]'s implementation to allow for more practical matrix multiplication by Cohn & Uman' method.

On the other hand, the author of this research wishes to clarify the discrepancy (on how the logic follows to generalize the initial claim) described in Section 5.1 regarding the Coppersmith and Winograd identity. Moreover, since no research has been able to rule out the possibility of obtaining $\omega = 2$ using Cohn & Umans method, continual research into this area may produce a better bound on ω on the current best known bound, not limited by any known bounds for other methods.

This also points out some limitation in our current research, only relying on zeroing out a group tensor instead of using stronger (monomial) degeneration. This lead to excessively large group G, which does yield a fast matrix multiplication algorithm, but is unsuitable for multiplying matrices of usual sizes. Therefore, a way to allow our current research to catch up with current research is to figure out how to formulate the "triple product property" to generate the group tensor into a better matrix multiplication tensor. This also showcases our limitation of focusing on generating practical matrix multiplication algorithm instead of pursuing theoretical bounds in previous research.

References

[1] Charu C. Aggarwal. *Linear Algebra and Optimization for Machine Learning [electronic resource] : A Textbook / by Charu C. Aggarwal.* eng. 1st ed. 2020. Cham: Springer International Publishing, 2020. ISBN: 3-030-40344-0.

[2] Josh Alman. "Limits on the Universal Method for Matrix Multiplication". In: *arvix e-prints* (2019). URL: `https://arxiv.org/abs/1812.08731`.

[3] Josh Alman and Virginia Vassilevska Williams. "Limits on All Known (and Some Unknown) Approaches to Matrix Multiplication". In: *Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation*. ISSAC '19. Beijing, China: Association for Computing Machinery, 2019,

p. 10. ISBN: 9781450360845. DOI: 10.1145/3326229.3326282. URL: https://doi.org/10.1145/3326229.3326282.

[4] Josh Alman and Virginia Vassilevska Williams. "Further Limitations of the Known Approaches for Matrix Multiplication". In: *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Ed. by Anna R. Karlin. Vol. 94. Dagstuhl, Germany, 2018, 25:1–25:15. ISBN: 978-3-95977-060-6. DOI: 10.4230/LIPIcs.ITCS.2018.25.

[5] Andris Ambainis and Yuval Filmus. *On the Coppersmith – Winograd method*. 2014. URL: https://api.semanticscholar.org/CorpusID:17684930.

[6] Andris Ambainis, Yuval Filmus, and François Le Gall. "Fast Matrix Multiplication: Limitations of the Coppersmith-Winograd Method". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing* (2014). URL: https://api.semanticscholar.org/CorpusID:8332797.

[7] Matthew Anderson and Siddharth Barman. *The Coppersmith-Winograd Matrix Multiplication Algorithm*. 2009. URL: https://www-auth.cs.wisc.edu/lists/theory-reading/2009-December/pdfmN6UVeUiJ3.pdf.

[8] Matthew Anderson, Zongliang Ji, and Anthony Yang Xu. "Matrix Multiplication: Verifying Strong Uniquely Solvable Puzzles". In: *Theory and Applications of Satisfiability Testing – SAT 2020*. Ed. by Luca Pulina and Martina Seidl. Cham: Springer International Publishing, 2020, pp. 464–480. ISBN: 978-3-030-51825-7.

[9] Matthew Anderson and Vu Le. "Efficiently-Verifiable Strong Uniquely Solvable Puzzles and Matrix Multiplication". In: *Computing and Combinatorics - 29th International Conference, COCOON 2023, Hawaii, HI, USA, December 15-17, 2023, Proceedings, Part II*. Ed. by Weili Wu and Guangmo Tong. Vol. 14423. Lecture Notes in Computer Science. Springer, 2023, pp. 41–54. DOI: 10.1007/978-3-031-49193-1_4. URL: https://doi.org/10.1007/978-3-031-49193-1%5C_4.

[10] Jonah Blasiak et al. "On cap sets and the group-theoretic approach to matrix multiplication". In: *Discrete Anaysis* (Aug. 2023). DOI: 10.19086/da.1245.

[11] Jonah Blasiak et al. "Which groups are amenable to proving exponent two for matrix multiplication?" In: *ArXiv* abs/1712.02302 (2017). URL: https://api.semanticscholar.org/CorpusID:24486404.

[12] Peter Bürgisser, Michael Clausen, and Mohammad A. Shokrollahi. *Algebraic Complexity Theory*. 1st. Springer Publishing Company, Incorporated, 2010. ISBN: 3642082289.

[13] Henry Cohn and Christopher Umans. “A group-theoretic approach to fast matrix multiplication”. In: *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* (2003), pp. 438–449.

[14] Henry Cohn et al. “Group-theoretic algorithms for matrix multiplication”. In: *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)* (2005), pp. 379–388.

[15] Don Coppersmith and Shmuel Winograd. “Matrix multiplication via arithmetic progressions”. In: *Journal of Symbolic Computation* 9.3 (1990). Computational algebraic complexity editorial, pp. 251–280. ISSN: 0747-7171. DOI: https://doi.org/10.1016/S0747-7171(08)80013-2. URL: https://www.sciencedirect.com/science/article/pii/S0747717108800132.

[16] Thomas H. Cormen et al. *Introduction to Algorithms*. 2nd ed. The MIT Press, 2001.

[17] Khai Dong. *Cohn-Umans Matrix Multiplication and Wedderburn Decomposition*. 2023.

[18] Ran Duan, Hongxun Wu, and Renfei Zhou. “Faster Matrix Multiplication via Asymmetric Hashing”. In: *arXiv e-prints*, arXiv:2210.10173 (Oct. 2022), arXiv:2210.10173. DOI: 10.48550/arXiv.2210.10173. arXiv: 2210.10173 [cs.DS].

[19] Zachary Dubinsky. *Fast Matrix Multiplication and the Wedderburn-Artin Theorem*. 2023.

[20] David S. Dummit and Richard M. Foote. *Abstract Algebra*. en. 3rd ed. Nashville, TN: John Wiley Sons, 2003. ISBN: 9780471433347.

[21] Johan Håstad. “Tensor rank is NP-complete”. In: *Journal of Algorithms* 11.4 (1990), pp. 644–654. ISSN: 0196-6774. DOI: https://doi.org/10.1016/0196-6774(90)90014-6. URL: https://www.sciencedirect.com/science/article/pii/0196677490900146.

[22] François Le Gall. “Powers of tensors and fast matrix multiplication”. In: *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*. ISSAC ’14. Kobe, Japan: Association for Computing Machinery, 2014, pp. 296–303. ISBN: 9781450325011. DOI: 10.1145/2608628.2608664. URL: https://doi.org/10.1145/2608628.2608664.

[23] Will Sawin. “Bounds for Matchings in Nonabelian Groups”. In: *Electron. J. Comb.* 25 (2017), p. 4. URL: https://api.semanticscholar.org/CorpusID:54446488.

[24] V. Strassen. “Gaussian Elimination is not Optimal.” In: *Numerische Mathematik* 13 (1969), pp. 354–356. URL: http://eudml.org/doc/131927.

[25] V. Strassen. “The asymptotic spectrum of tensors and the exponent of matrix multiplication”. In: *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*. 1986, pp. 49–54. DOI: 10.1109/SFCS.1986.52.

[26] Virginia Vassilevska Williams et al. "New Bounds for Matrix Multiplication: from Alpha to Omega". In: *arXiv e-prints*, arXiv:2307.07970 (July 2023), arXiv:2307.07970. DOI: 10.48550/arXiv.2307.07970. arXiv: 2307.07970 [cs.DS].

[27] S. Winograd. "On multiplication of 2 × 2 matrices". In: *Linear Algebra and its Applications* 4.4 (1971), pp. 381–388. ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(71)90009-7. URL: https://www.sciencedirect.com/science/article/pii/0024379571900097.