KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Mathematics

Second cycle 30 credits

Tensor Rank and Support Rank in the Context of Algebraic Complexity Theory

PELLE ANDERSSON

Stockholm, Sweden 2024

Abstract

Starting with the work of Volker Strassen, algorithms for matrix multiplication have been developed which are time complexity-wise more efficient than the standard algorithm from the definition of multiplication. The general method of the developments has been viewing the bilinear mapping that matrix multiplication is as a three-dimensional tensor, where there is an exact correspondence between time complexity of the multiplication algorithm and tensor rank. The latter can be seen as a generalisation of matrix rank, being the minimum number of terms a tensor can be decomposed as. However, in contrast to matrix rank there is no general method of computing tensor ranks, with many values being unknown for important three-dimensional tensors. To further improve the theoretical bounds of the time complexity of matrix multiplication, support rank of tensors has been introduced, which is the lowest rank of tensors with the same support in some basis. The goal of this master's thesis has been to go through the history of faster matrix multiplication, as well as specifically examining the properties of support rank for general tensors. In regards to the latter, a complete classification of rank structures of support classes is made for the smallest non-degenerate tensor product space in three dimensions. From this, the size of a support can be seen affecting the pool of possible ranks within a support class. At the same time, there is in general no symmetry with regards to support size occurring in the rank structures of the support classes, despite there existing a symmetry and bijection between mirrored supports. Discussions about how to classify support rank structures for larger tensor product spaces are also included.

Contents

I would like to give special thanks to my supervisor Mats Boij and my family for their invaluable help and support throughout this degree project.

1 Introduction

1.1 Beginnings: Strassen's algorithm

During the 1960's, the first known improvement, time complexity-wise, to matrix multiplication was discovered. When Volker Strassen tried to prove Gaussian elimination was optimal, which includes showing that the common method of matrix multiplication is optimal, he inadvertently found that at the cost of more additions a 2×2-matrix product could be written with seven products instead of the usual eight.[1] Using notation

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \; B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \; AB = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

the product AB can, by definition, be written using a total of $2^3 = 8$ multiplications of matrix entries as

$$C_{ij} = \sum_{k=1}^{2} A_{ik} B_{kj}.$$

But the same result can be achieved using the seven multiplications

$$\begin{aligned} \mathrm{I} &= (A_{11} + A_{22})(B_{11} + B_{22}), \\ \mathrm{II} &= (A_{21} + A_{22}) B_{11}, \\ \mathrm{III} &= A_{11}(B_{12} - B_{22}), \\ \mathrm{IV} &= A_{22}(-B_{11} + B_{21}), \\ \mathrm{V} &= (A_{11} + A_{12}) B_{22}, \\ \mathrm{VI} &= (-A_{11} + A_{21})(B_{11} + B_{12}), \\ \mathrm{VII} &= (A_{12} - A_{22})(B_{21} + B_{22}) \end{aligned} \tag{1}$$

and adding them to

$$\begin{aligned} C_{11} &= \mathrm{I} + \mathrm{IV} - \mathrm{V} + \mathrm{VII}, \\ C_{12} &= \mathrm{III} + \mathrm{V}, \\ C_{21} &= \mathrm{II} + \mathrm{IV}, \\ C_{22} &= \mathrm{I} + \mathrm{III} - \mathrm{II} + \mathrm{VI}. \end{aligned} \tag{2}$$

If the entries of matrices A, B are relatively small numbers, addition and multiplication of the entries take roughly the same amount of time, which means this new algorithm is not an improvement over the common one. However, if A, B are block matrices, then addition is often faster. If the blocks are of size $m \times m$ then addition has a time complexity of $\mathcal{O}(m^2)$ compared to multiplication having time complexity $\mathcal{O}(m^3)$. If A, B are of size $n \times n$ and $n = 2^k$, then the seven product algorithm above can be used recursively for each multiplication, which is commonly called *Strassen's algorithm.*

The assumption $n = 2^k$ is always possible since one otherwise can construct matrices A' and B' of size $2^k \times 2^k$ for the smallest k such that $n \leq 2^k$ where $A'_{ij} = A_{ij}$ for $i, j \leq n$ and $A'_{ij} = 0$ otherwise, and similarly for B' and B. Then $A'B'$ will be equal to $(AB)'$, defined analogous to A' and B', which corresponds exactly to AB, and results in an algorithm for multiplying $n \times n$-matrices using only multiplication of $2^k \times 2^k$-matrices.

Assuming operations between scalars taking $\mathcal{O}(1)$ time, Strassen's algorithm has a time complexity of $\mathcal{O}\left(n^{\log_8 7}\right) \approx \mathcal{O}\left(n^{2.8}\right)$, which will be shown later in Section 3.1. This result, published in Volker Strassen's article "Gaussian Elimination is not Optimal" in 1969 [2], became the starting point for faster matrix multiplication, but also led to creating the field of algebraic complexity theory in general. The core of the latter is translating algebra-related problems in computer science and complexity theory into pure algebra. This can be done to the time complexity of matrix multiplication by viewing the multiplication algorithm as a bilinear map, and viewing that map as an element in a tensor product of three vector spaces. Finally, this tensor element will have an algebraic property known as tensor rank, roughly defined as the minimum number of terms the element can be written as. This property of rank will, with respect to matrix size, asymptotically correspond exactly to the time complexity of the multiplication algorithm.

Since then, further relaxations of tensor rank has been introduced with the primary purpose at first being to further bound the time complexity of matrix multiplication. These include border rank and support rank of a tensor, which focus on properties of approximations and the support in a particular basis, respectively, of that particular tensor. Support of a tensor will be the set of basis elements where the tensor has non-zero coefficients, and different tensors with same support can be interpreted as the same bilinear algorithm but with different weights for each coordinate vector. While introduced as tools for studying matrix multiplication, these newer rank definitions can be used for all tensors and multilinear algorithms, and are in of themselves worthy of study.

1.2 Purpose and outline

This text will focus both on the history of the time complexity of matrix multiplication, as well as working out properties of tensor rank and support rank.

Section 2 will introduce the correspondence between elements of tensor products and bilinear algorithms, including the special case of matrix multiplication. Tensor rank - the most important definition of the text - will also be defined here. In Section 3, the correspondence between tensor rank and time complexity of matrix multiplication is derived, and the history and recent developments of bounding the time complexity are discussed. In tandem with the latter, the new relaxations of rank border rank and support rank are also introduced. Then, in Section 4, the structures within sets of tensors with the same support is analyzed. The main issue here is the tensor rank structure, and the section will include a complete classification of rank structures of support classes for the smallest true, non-degenerate three-dimensional tensor product. The text ends with a discussion in Section 5.

1.3 Clarifications

Throughout the text, some important derived results will be summarized in propositions, but will not have a formal proof directly after. Instead, the formal arguments and the derivation of the results are found before the actual formulation of the proposition. Italic text will, except when mentioning other written works, consistently be reserved for mathematical definitions.

1.3.1 Prerequisites

The reader is assumed to be familiar with elementary tensor theory found in an undergraduate course in linear algebra, and the concept of time complexity within computer science. Other preliminaries are groups and basic Galois theory, such as the concepts field extensions and characteristics of fields, which can for example be referred to in the book *Abstract Algebra: Theory and Applications* by Judson[3]. The concepts of the prerequisite theories will be taken as granted and not be defined in this text, and most results and theorems from these will not be cited directly.

1.3.2 Other clarifications and notation

- Intervals of integers are denoted $[a, b] = \{a, a + 1, \ldots, b\}$ and $[n] = \{1, 2, \ldots, n\}$.
- Here, vector spaces have finite dimension and can be over any field K.
- A *three-dimensional tensor*, or just tensor, refers to an element in a tensor product $X \otimes Y \otimes Z$ of three vector spaces X, Y, Z. When $X \otimes Y \otimes Z$ has a basis $\{x_i \otimes y_j \otimes z_k\}_{i,j,k}$ where x_i, y_j, z_k are vectors in X, Y, Z respectively, $x_i y_j z_k$ is sometimes used as shorthand notation for the tensor $x_i \otimes y_j \otimes z_k$ in order to save space.
- Ordo-notation is used for sets of functions, with the common definition

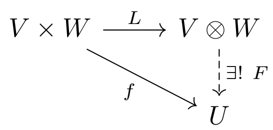
  $$\mathcal{O}(g) := \{f : \exists N, C \ , \ f(n) \leq Cg(n) \, , \ \forall n \geq N\},$$

  and sums with ordo-notation are denoted $f + \mathcal{O}(g) = \{f + f' : f' \in \mathcal{O}(g)\}$.

2 Tensor perspective

2.1 Matrix multiplication tensors

First off, by the universal property of tensor product, each bilinear map $f : V \times W \to U$ between vector spaces has a unique factorization



where the bilinear map L is defined by $L(x,y) = x \otimes y$, and F is linear. Additionally, there is a natural isomorphism between spaces $V^* \otimes U \cong \hom(V,U)$, which means for each bilinear map $f : V \times W \to U$ there is a unique tensor element in $(V \otimes W)^* \otimes U \cong V^* \otimes W^* \otimes U$. In the case of a matrix multiplication map

$$f : M_{a\times b} \times M_{b\times c} \longrightarrow M_{a\times c},$$
$$(A,B) \longmapsto AB$$

where $M_{a\times b}$ is the vector space of $a \times b$-matrices, the corresponding tensor

$$t(f) \in M^*_{a\times b} \otimes M^*_{b\times c} \otimes M_{a\times c}$$

can be found in the following way.

First, choose bases $\{x_{ij}\}, \{y_{kl}\}, \{z_{mn}\}$ for $M_{a\times b}$, $M_{b\times c}$ and $M_{a\times c}$ respectively, where $x_{ij} \in M_{a\times b}$ is the matrix with entries $(x_{ij})_{i'j'} = \delta_{ii'}\delta_{jj'}$, and similarly for y_{kl}, z_{mn} and the other matrix spaces, and let $\{x^*_{ij}\}, \{y^*_{kl}\}$ be the corresponding dual bases such that $x^*_{ij}(x_{i'j'}) = \delta_{ii'}\delta_{jj'}$. Then there is a basis $\{x^*_{ij} \otimes y^*_{kl} \otimes z_{mn}\}$ for $M^*_{a\times b} \otimes M^*_{b\times c} \otimes M_{a\times c}$ where $t(f)$ is written as

$$t(f) = \sum_{i,j,k,l,m,n} t_{ijklmn} x^*_{ij} \otimes y^*_{kl} \otimes z_{mn},$$

where the coefficients are determined by how f maps inputs (x_{ij}, y_{jk}). From the definition of matrix multiplication, the (i,j):th entry in $f(x_{i'j'}, y_{j'k'})$ will be

$$\sum_k (x_{i'j'})_{ik} (y_{k'l'})_{kj} = \sum_k \delta_{i'i}\delta_{j'k}\delta_{k'k}\delta_{l'j} = \delta_{i'i}\delta_{l'j}\delta_{j'k'},$$

and so the entire output matrix can be written as

$$f(x_{i'j'}, y_{j'k'}) = \sum_{i,j} \left(f(x_{i'j'}, y_{j',k'})\right)_{ij} z_{ij} = \sum_{i,j} \delta_{i'i}\delta_{l'j}\delta_{j'k'} z_{ij} = \delta_{j'k'} z_{i'l'}.$$

Since f and $t(f)$ correspond by

$$f(A,B) = \sum_{i,j,k,l,m,n} t_{ijklmn} x^*_{ij}(A) \otimes y^*_{kl}(B) \otimes z_{mn},$$

putting $(x_{i'j'}, y_{j'k'})$ into f and $t(f)$ results in

$$\delta_{j'k'} z_{i'l'} = \sum_{i,j,k,l,m,n} t_{ijklmn} \delta_{ii'}\delta_{jj'}\delta_{kk'}\delta_{ll'} z_{mn}.$$

Both sides of the equation are respectively equal to

$$\sum_{m,n} \delta_{i'm}\delta_{l'n}\delta_{j'k'} z_{mn} = \sum_{m,n} t_{i'j'k'l'mn} z_{mn},$$

and so the coefficients can be identified as $t_{ijklmn} = \delta_{im}\delta_{ln}\delta_{jk}$. Thus, the entire tensor becomes, after swapping k and l with each other in order to end up with only indices i, j, k,

$$t(f) = \sum_{i,j,k} x^*_{ij} \otimes y^*_{jk} \otimes z_{ik}.$$

This result can be simplified by $t(f)$ corresponding exactly to a tensor

$$\langle a, b, c \rangle := \sum_{\substack{i \in [1,a] \\ j \in [1,b] \\ k \in [1,c]}} x_{ij} \otimes y_{jk} \otimes z_{ki} \tag{3}$$

in $M_{a\times b} \otimes M_{b\times c} \otimes M_{c\times a}$. This is chosen in order to get more symmetry, where dual spaces are isomorphic to each other if they have finite dimension. This will from now on be the tensor of choice that corresponds to matrix multiplication $M_{a\times b} \times M_{b\times c} \longrightarrow M_{a\times c}$. It should however always be kept in mind that $\langle a, b, c \rangle$ has been achieved using transposition of matrices and moving from dual spaces. This will then need to be accounted for when for example doing a change of bases for $M_{a\times b}, M_{b\times c}, M_{a\times c}$, though such changes will not occur throughout this text.

Moving on to describing the time complexity using algebraic language, a tensor on the form

$$t = \sum_{i=1}^{N} u_i \otimes v_i \otimes w_i \in X \otimes Y \otimes Z,$$

where u_i, v_i, w_i can be any vectors in their respective spaces, will correspond to a bilinear map where the u_i and v_i:s take inputs, and the w_i:s describes the resulting output in Z. Then the time to input some (x, y) into each term of t will be proportional to N, which means that the minimal number of terms that t can be written as must be asymptotically bounded from above by the time complexity of an optimal algorithm for the bilinear map. For $t = \langle n, n, n \rangle$, the bound will hold in the opposite direction as well, and so there is an equivalence between time complexity and the minimum number of terms of a tensor, the latter of which is called the rank of a tensor.

2.2 Tensor rank

More specifically, a three-dimensional tensor $t \in X \otimes Y \otimes Z$ is said to have *rank one* if there are $(x, y, z) \in X \times Y \times Z$, all non-zero, such that $t = x \otimes y \otimes z$. Alternatively, if one has bases $\{x_i\}, \{y_j\}, \{z_k\}$ for the spaces X, Y and Z, then a non-zero t has rank one if there are scalars a_i, b_j, c_k such that

$$t = \left(\sum_i a_i x_i\right) \otimes \left(\sum_j b_j y_j\right) \otimes \left(\sum_k c_k z_k\right) = \sum_{i,j,k} a_i b_j c_k x_i \otimes y_j \otimes z_k. \tag{4}$$

The *rank* of a three-dimensional tensor t, denoted $R(t)$ is then defined as the smallest number of terms with rank one that t can be written as a sum of. The zero element is defined to have rank equaling zero.

A basic property of rank is that it is preserved under *tensor products of linear maps*

$$A \otimes B \otimes C : X' \otimes Y' \otimes Z' \longrightarrow X \otimes Y \otimes Z,$$

which are constructed using linear maps

$$A : X \to X',\ B : Y \to Y',\ C : Z \to Z',$$

where $(A \otimes B \otimes C)(x \otimes y \otimes z) = A(x) \otimes B(y) \otimes C(z)$ and acts linearly. Then a rank one tensor is mapped to either another rank one tensor or the zero tensor, and so a product of linear maps cannot increase the rank of a tensor. In particular, this means that rank is basis-invariant, since a change of bases for X, Y and Z is a tensor product of linear maps, and so every rank of t in each trio of bases $\{x_i\}, \{y_j\}, \{z_k\}$ bound each other, therefore making them all equal. A tensor $s \in X' \otimes Y' \otimes Z'$ is said to be a *restriction* of $t \in X \otimes Y \otimes Z$ if there is a tensor product of linear maps $A \otimes B \otimes C : X' \otimes Y' \otimes Z' \to X \otimes Y \otimes Z$ such that $s = (A \otimes B \otimes C)\, t$, which is denoted $s \leq t$. Another term for this is that s is *embedded* in t, with the intuition that "s lies inside of t".

The results of rank can be summarized as follows.

Proposition 1. *If s and t are three-dimensional tensors where s is a restriction of t using maps $A \otimes B \otimes C$, then $R(s) \leq R(t)$, where equality holds if A, B, C are bijective. In particular, rank in $X \otimes Y \otimes Z$ is independent of choice of bases for X, Y, Z.*

If one of the the spaces X, Y, Z has dimension 1, then $X \otimes Y \otimes Z$ will have the same tensor structure as a two-dimensional $U \otimes W$, which as seen before is naturally isomorphic to linear maps between U and W. This gives a definition of rank for two-dimensional tensors, which will correspond exactly to rank for linear maps and matrices. The matrix rank and three-dimensional rank differ in that for the former there is the Fundamental theorem of linear algebra categorizing maps with maximum rank, and matrix ranks can be computed in polynomial time with respect to matrix size by using Gaussian elimination. In general there is no similar categorization for three-dimensional rank, and computation of rank given a tensor is an NP-hard problem[4], and thus highly unlikely to have polynomial time complexity. As such, ranks of three-dimensional tensors are seldom computed exactly, with the most straight-forward way of bounding a rank is to explicitly write a tensor as a decomposition of rank one terms.

2.2.1 Largest rank in a tensor product

Directly from the definition, the maximum rank within a tensor product $X \otimes Y \otimes Z$ cannot be larger than the dimension of it as a vector space, but this can be narrowed down, even though the exact value for general tensor products is as of yet unknown. Say that the dimensions are $\dim X = n_1$, $\dim Y = n_2$, and $\dim Z = n_3$. Any tensor $t = \sum_{i,j,k} t_{ijk} x_i y_j z_k$ can be decomposed into so-called *slices*, where an x_a-*slice* is the two-dimensional tensor

$$t_{x_a} = \sum_{j,k} t_{ajk} y_j \otimes z_k$$

and similarly for y_b- and z_c-slices. The decomposition of t into slices is then

$$t = \sum_{i=1}^{n_1} x_i \otimes t_{x_i}. \tag{5}$$

But since each slice t_{x_i} is a two-dimensional tensor in $Y \otimes Z$, their ranks are at most $\min\{n_2, n_3\}$, and so (5) can be decomposed into $l \cdot \min\{n_2, n_3\}$ rank one tensors. However, choosing to decompose using x-slices is arbitrary, and so using a similar method a decomposition into $n_i n_j$ rank one tensors can be made. The maximum rank $r_{\max}$ in $X \otimes Y \otimes Z$ is thus at most

$$r_{\max} \leq \min_{i,j \in \{1,2,3\}} n_i n_j.$$

The maximum rank can also be bounded from below where the technique differs whether the underlying field is finite or not. If the field K is finite and of size $|K| = k$, then the number of tensors of rank at most one, denoted $N(1)$, can be used to roughly bound from above the number of tensors of rank at most r, denoted $N(r)$, by $N(r) \leq N(1)^r$ since one chooses r different rank one-tensors. A approximation for the maximum rank can then be obtained by comparing $N(r)$ with the number of tensors in $X \otimes Y \otimes Z$, namely $k^{n_1 n_2 n_3}$. At first glance, it holds that $N(1)$ is at most $k^{n_1+n_2+n_3}$ since a tensor of rank at most one is of the form $t_1 = u \otimes v \otimes w$. However, there are ways of writing the same t_1 using different vectors u, v, w. For each choice of u, v, w, there are $(k-1)^2$ choices for non-zero elements $a, b \in K$ such that $t_1 = (abu) \otimes (a^{-1}v) \otimes (b^{-1}w)$. This then means that $N(1) \leq k^{n_1+n_2+n_3}/(k-1)^2$. As such, the number of all tensors in $X \otimes Y \otimes Z$, that then have rank at most $r_{\max}$, is bounded by

$$k^{n_1 n_2 n_3} = N(r_{\max}) \leq \left(\frac{k^{n_1+n_2+n_3}}{(k-1)^2} \right)^{r_{\max}}$$

which means the maximum rank can be bounded according to

$$r_{\max} \geq \frac{n_1 n_2 n_3}{n_1 + n_2 + n_3 - 2\log_k(k-1)}.$$

In particular,

$$r_{\max} \geq \frac{n_1 n_2 n_3}{n_1 + n_2 + n_3} \tag{6}$$

holds for all finite fields.

For general fields K one can use a dimensional argument instead of counting the number of tensors. All tensors in $X \otimes Y \otimes Z$ can be written as

$$t = \sum_{i=1}^{r_{\max}} u_i \otimes v_i \otimes w_i \tag{7}$$

where u_i, v_i, w_i are any vectors in X, Y, Z. This means that one has $r_{\max} \cdot (n_1 + n_2 + n_3)$ degrees of freedom of choosing elements of K when constructing any tensor t. But $X \otimes Y \otimes Z$ is a vector space of dimension $n_1 n_2 n_3$, and the degrees of freedom cannot be less than the that. As such, (6) holds for all fields.

It should be noted that $\log_k(k-1)$ converges to 1 as k increases, which begs the question if

$$r_{\max} \geq \frac{n_1 n_2 n_3}{n_1 + n_2 + n_3 - 2} \tag{8}$$

for infinite fields. This can be seen to hold when for example $K = \mathbb{R}$, where the following the argument above one has $r_{\max} \cdot (n_1 + n_2 - 1 + n_3 - 1)$ degrees of freedom when constructing any real tensor. This, since all the vectors v_i and w_i can be assumed to be non-zero and $|v_i| = 1 = |w_i|$ in (7), as their norms can be factored into u_i instead. This means that there are $n_2 - 1$ and $n_3 - 1$ degrees of freedom when choosing v_i and w_i respectively, since they are points on the projective spaces of Y and Z. It is beyond the scope of this text, but it was proved by Howell in 1978 [5] that (8) actually holds for all infinite fields.

The results can be summarized as follows.

Proposition 2. *The maximum tensor rank $r_{\max}$ in a tensor product $X \otimes Y \otimes Z$, with dimensions $\dim X = n_1$, $\dim Y = n_2$, $\dim Z = n_3$ and any underlying field K, can be bounded by*

$$\frac{n_1 n_2 n_3}{n_1 + n_2 + n_3} \leq r_{\max} \leq \min_{i,j \in \{1,2,3\}} n_i n_j .$$

Here, the lower bound becomes tighter the larger the field gets. This should not be interpreted as the maximal rank becoming larger for larger fields, as $r_{\max}$ is still unknown for general dimensions. As of yet, the lower bound is the same as when it was computed in 1978, at least when considering general fields and dimensions. In contrast, finding tighter upper bounds have generally been more fruitful, where there is an upper bound of $n^2 - n - 1$ for $K^n \otimes K^n \otimes K^n$ when $\operatorname{char} K = 0$.[6]

Regardless, it still holds for when X, Y, Z have the same dimension n that the maximum rank is of size $\mathcal{O}(n^2)$, in comparison to maximum matrix rank in dimension n being n. All the arguments done within this text when approximating the maximum rank can be applied to ranks of higher tensor products $X_1 \otimes X_2 \otimes \dots X_d$, that is, they do not utilise the unique properties of three-dimensional tensor products. The rank of a d-dimensional tensor is defined analogously to that of three dimensions, and the result is that a maximal d-dimensional tensor rank is of magnitude $\mathcal{O}(n^{d-1})$ when $\dim X_i = n$ for all i.

2.3 Rank of matrix multiplication tensors

For the case of the tensor $\langle 2, 2, 2 \rangle$, it has rank $\leq 2^3 = 8$ from (3), but can be bounded by 7 from Strassen's algorithm. First, construct seven two-dimensional tensors s_a where each corresponds to a equation in (1) by exchanging all A_{ij} with x_{ij} and all B_{kl} with y_{kl}, using same variables as in (3), and use tensor multiplication $\otimes$ instead of multiplications. For example,

$$\begin{aligned} s_{\mathrm{I}} &= (x_{11} + x_{22}) \otimes (y_{11} + y_{22}) , \\ s_{\mathrm{II}} &= (x_{21} + x_{22}) \otimes y_{11}, \end{aligned}$$

and so on. Then following the recipe in (2), by multiplying the s_a:s with z-matrices and linearly combining the terms, results in

$$\begin{aligned}\langle 2,2,2\rangle = &(s_{\mathrm{I}} + s_{\mathrm{IV}} - s_{\mathrm{V}} + s_{\mathrm{VII}}) \otimes z_{11} + (s_{\mathrm{II}} + s_{\mathrm{IV}}) \otimes z_{12} + \\ &+ (s_{\mathrm{III}} + s_{\mathrm{V}}) \otimes z_{21} + (s_{\mathrm{I}} + s_{\mathrm{III}} - s_{\mathrm{II}} + s_{\mathrm{VI}}) \otimes z_{22},\end{aligned} \tag{9}$$

remembering the transpositions made to the z-vectors in (3). Using the s_a:s as common factors instead of the z:s, this can be decomposed as

$$\begin{aligned}\langle 2,2,2\rangle = &s_{\mathrm{I}} \otimes (z_{11} + z_{22}) + s_{\mathrm{II}} \otimes (z_{12} - z_{22}) + s_{\mathrm{III}} \otimes (z_{21} + z_{22}) + s_{\mathrm{IV}} \otimes (z_{11} + z_{12}) + \\ &+ s_{\mathrm{V}} \otimes (-z_{11} + z_{21}) + s_{\mathrm{VI}} \otimes z_{22} + s_{\mathrm{VII}} \otimes z_{11},\end{aligned}$$

which is made out of seven rank one tensors, thus bounding $R(\langle 2,2,2\rangle) \leq 7$. The same thinking applies given an algorithm for matrices of other sizes using N multiplication, resulting in a decomposition of the matrix multiplication tensor into N tensors of rank one.

This also works in the other direction. Given a decomposition of a matrix multiplication tensor $\langle l,m,n\rangle$ into N rank one terms, there are exactly N different two-dimensional factors involving x, y-variables, which corresponds to N multiplications similarly to (1). Also, decomposing $\langle l,m,n\rangle$ using z:s as common factors, as in (9), gives a recipe similarly to (2) how the matrix elements of the product AB are constructed using the N multiplications. This correspondence exemplifies one of the most important notions to keep in mind when moving from bilinear algorithms to tensors. Although the matrix multiplication tensors are purely algebraic objects, where the correspondence to the bilinear algorithms can essentially be forgotten, any manipulations made to these tensors will always have some implicit matrix algorithm behind them.

2.3.1 Attempts at explaining Strassen's algorithm

As stated before, the multiplication algorithm (1) and (2) was discovered inadvertently by Strassen, and it may at first glance seem as an arbitrary algorithm. It is for example non-symmetric in comparison to the highly symmetrical usual method of matrix multiplication, it was never explained how one would arrive at the algorithm[1][2], and there is no mention that this is the only method which induces a decomposition of the corresponding tensor $\langle 2,2,2\rangle$ into seven rank one terms. Since then, there have been attempts at explaining Strassen's algorithm, insofar as to give a proof of the existence of a rank seven decomposition of $\langle 2,2,2\rangle$ without giving an explicit example of such a decomposition, and then derive Strassen's decomposition from there[1].

Grochow and Moore found, inspired by the symmetry of matrix multiplication, a symmetry known as *unitary 2-design* where a set of n-dimensional vectors $S \subset K^n$ fulfill

$$\sum_{v\in S} v = 0 \quad \text{and} \quad \frac{1}{|S|}\sum_{v\in S} v^* \otimes v = \frac{1}{n}\,\mathrm{Id}$$

where Id is the tensor which corresponds to the identity map $K^n \to K^n$, and $v^* \in (K^n)^*$ is made up of the same coefficients as v in a basis $\{e_1, \ldots, e_n\}$ for K^n, but in a dual basis $\{e_1^*, \ldots, e_n^*\}$ for $(K^n)^*$ instead. Slightly different language is used by Grochow and Moore than in this text, opting for matrices described by Dirac notation for outer products, but the results are similar.[1]

It can then be shown that $\langle n, n, n\rangle$, or more specifically an equivalent tensor, can be decomposed using vectors from a unitary 2-design $S = \{v_1, \ldots, v_s\}$ as

$$\mathrm{Id} \otimes \mathrm{Id} \otimes \mathrm{Id} + \frac{n^3}{s^3} \cdot \sum_{i,j,k \text{ distinct}} ((v_j - v_i)^* \otimes v_i) \otimes ((v_k - v_j)^* \otimes v_j) \otimes ((v_i - v_k)^* \otimes v_k).$$

This then means that the rank of $\langle n, n, n\rangle$ is bounded from above by $1 + |S|(|S| - 1)(|S| - 2)$. It can then be checked that, for $n = 2$, $S = \{(1, 0),\ (-1/2, \sqrt{3}/2),\ (-1/2, -\sqrt{3}/2)\} \subset \mathbb{C}^2$ is a unitary 2-design, inducing a decomposition into $1 + 3 \cdot 2 \cdot 1 = 7$ rank one terms.[1] This decomposition is not the same as Strassen's, but it has been shown by De Groote that, specifically for $n = 2$, optimal decompositions of $\langle 2, 2, 2\rangle$ are equivalent up to using invertible linear maps A, B, C on K^2 according to

$$\langle 2, 2, 2\rangle = (A \otimes B \otimes C)\langle 2, 2, 2\rangle(A^{-1} \otimes B^{-1} \otimes C^{-1})$$

which corresponds to a change of bases for the matrix spaces.[7]

So from this, one can arrive at Strassen's algorithm for $\langle 2, 2, 2\rangle$. However, it does not explain how one would arrive at a decomposition which works for all fields K, or indeed all rings with multiplicative identity, which Strassen's does, as it uses only elements 0 and 1, and ring operations between them. For instance, the unitary 2-design S from above does not work in characteristic two or three, and it is still an open problem to find unitary 2-designs for all fields and rings, or in any other way explain why Strassen's algorithm can be applied to all rings[1].

2.4 Method of substitution

There are other ways to bound ranks without finding an explicit decomposition, with one example being to find lower bounds using the so-called method of substitution. The idea is to view a tensor as a function of variables x_i, y_j and z_k, and then manipulate the tensor by substituting one $x_{i'}$ as a linear combination of the other x_i:s, usually the goal being to turn specific rank one terms to zero. This can for example be applied to the tensor $t = x_0y_0z_0 + x_0y_1z_1 + x_1y_0z_1$ within dimension 2^3, which does not have rank equal to one. For a proposed decomposition of rank two, $t = u_0 \otimes v_0 \otimes w_0 + u_1 \otimes v_1 \otimes w_1$, one of the terms must, as a function, depend on x_1, say the first one, which can always be made zero by substituting $x_1 = \alpha x_0$ for some $\alpha \neq 0$. This results in

$$t\left(x_0 = \alpha x_1\right) = u_1 \otimes v_1 \otimes w_1 = x_0y_0z_0 + x_0y_1z_1 + \alpha x_0y_0z_1 = x_0\left(y_0\left(z_0 + \alpha z_1\right) + y_1z_1\right).$$

Then, similarly to above, the y:s can be substituted according to $y_1 = \beta y_0$ to make the last term in this new tensor zero. This finally results in a zero tensor

$$t\left(x_1 = \alpha x_0, y_1 = \beta y_0\right) = 0 = x_0\left(y_0\left(z_0 + \alpha z_1\right) + \beta y_0z_1\right) = x_0y_0\left(z_0 + (\alpha + \beta)\, z_1\right).$$

But this last tensor clearly depends on z_0 and as such cannot be zero, resulting in a contradiction and that t must have rank larger than two, namely three.

Methods of substitution tend to only give very rough estimates of rank in the general case, however. In the relevant matrix multiplication tensor $\langle l, m, n\rangle = \sum_{i,j,k} x_{ij} \otimes y_{jk} \otimes z_{kj}$, with a decomposition of r rank one terms, one can make all but one of the terms zero, by doing $r-1$ substitutions. If $r \leq \max\{lm, mn, ln\}$, then this can be done using substitutions within only one of the three groups of variables, say $\{x_{ij}\}$, while still having one x_{ab} variable left not substituted. This results in a new rank one tensor

$$t_1 = x_{ab} \otimes \sum_{i,j,k} c_{ij} y_{jk} \otimes z_{ki}$$

for scalars c_{ij}, where in particular $c_{ab} = 1$. This is due to the fact that, in $\langle l, m, n\rangle$, x_{ab} is the only variable with the factor $\sum_k y_{bk} \otimes z_{ka}$, and so must stay unchanged when substituting the other x-variables. Since t_1 has rank one or smaller, it holds that

$$\sum_{i,j,k} c_{ij} y_{jk} \otimes z_{ki} = \sum_{i',j',k',l'} \alpha_{i'j'} \beta_{k'l'} y_{i'j'} \otimes z_{k'l'} \tag{10}$$

where $c_{ab} = 1$ implies $\alpha_{bk}\beta_{ka} = 1$ for all k. This should then mean that $\alpha_{b1}\beta_{2a} \neq 0$, which is the coefficient for $y_{a1} \otimes z_{2a}$. But looking at the left hand side of (10), this coefficient should be zero. This is a contradiction and so $R(\langle l, m, n\rangle)$ must be strictly larger than $\max\{lm, mn, ln\}$. Here, it was assumed that $n \geq 2$, but due to symmetry an assumption $\max\{l, m, n\} \geq 2$ works as well. It should be noted that

$$R(\langle l, m, n\rangle) > \max\{lm, mn, ln\}$$

only gives a rough estimate for the rank, being bounded from above by lmn.

3 Equivalence of rank and time complexity of matrix multiplication

3.1 Exponent of matrix multiplication

Define $\omega := \inf\{\beta : R(\langle n, n, n\rangle) \in \mathcal{O}(n^\beta)\}$ as the *exponent of matrix multiplication*, where so far the bounds $2 \leq \omega \leq 3$ are known from (3) and the method of substitution. Define $\tilde{\omega}$ the same way as ω except for replacing the rank with time complexity for $\langle n, n, n\rangle$'s corresponding bilinear algorithm, assuming operations between scalars taking constant time. Then $\tilde{\omega}$ should be interpreted as a description of the time complexity of matrix multiplication in general, and the goal is to show $\tilde{\omega} = \omega$.

The bound $\omega \leq \tilde{\omega}$ will hold for any tensor, not just ones corresponding to matrix multiplication. From the argument of correspondence between tensors and bilinear algorithms seen in Section 2.1, any bilinear algorithm defines a tensor decomposition

$$t = \sum_{i=1}^{N} u_i \otimes v_i \otimes w_i \in X \otimes Y \otimes Z,$$

which gives a method of computing the corresponding bilinear algorithm of t by inputting values from $X \times Y$ into u_i and v_i, and where then w_i describe the output in Z. Then the time of inputting and outputting for all terms will be proportional to N, and so the time complexity of the entire algorithm cannot be asymptotically faster than the minimal N, that is, the rank. In other words, $\omega \leq \tilde{\omega}$.

For the other bound, inspired by the proof of Bläser[8], block matrices and recursion will be applied. So, similarly to the case of Strassen's algorithm note that for any fixed $m > 1$, there is for each n a smallest integer k that $n \leq m^k$, filling remaining matrix entries with zeroes if equality does not hold, where the matrices are viewed as block matrices of size $m \times m$ with elements being $m^{k-1} \times m^{k-1}$ matrices. That the matrix size has been enlarged will not matter asymptotically in this case, which will be seen later. The goal is then to choose an m that results in a low rank $r = R(\langle m, m, m\rangle)$ in relation to ω since $\tilde{\omega}$ and ω are being compared. Given the rank r, there is, as seen in the previous section, an algorithm for $m \times m$-matrices using r multiplications and, say, c number of linear operations. While r and c are functions of m, the three are in this case all constants, and k is only the variable in relation to variation of n.

Then, applying recursion of block matrices to the $m \times m$-algorithm, the number of arithmetic operations of scalars $A(k)$ for multiplying $m^k \times m^k$ matrices, will be proportional to the time complexity corresponding to $\langle n, n, n\rangle$, and it will follow the recursion formula

$$A(k) \leq rA(k-1) + c \cdot \left(m^{k-1}\right)^2$$

with $A(0) = 1$. This can be rewritten as

$$A(k) \leq r^k cm^{2(k-1)} \sum_{i=0}^{k-2} \frac{r^i}{m^{2i}}$$

where the geometric sum can be written as

$$A(k) \leq r^k cm^{2(k-1)} \sum_{i=0}^{k-2} \frac{r^i}{m^{2i}} = \left(1 + \frac{cm^2}{r(r-m^2)}\right) r^k - \frac{cm^2}{r-m^2},$$

since $r > m^2$, as seen before when using method of substitution earlier. If so, and noting that in relation to variation of n the only variable is k, with m and of it dependent values r and c being constants, then $A(k) \in \mathcal{O}\left(r^k\right)$.

Now, comparing with starting matrix size n that may have been enlarged to m^k with $k = \lceil \log_m n \rceil$, it holds that $\mathcal{O}\left(r^{\lceil \log_m n \rceil}\right) = \mathcal{O}\left(r^{\log_m n}\right)$, that is, the rounding up does not matter asymptotically. The inclusion $\supset$ holds due to the rounded up function being larger, and for the other inclusion, it holds that

$$r^{\lceil \log_m n \rceil} \leq r^{1+\log_m n} = r \cdot r^{\log_m n}$$

for all n, and so $\mathcal{O}\left(r^{\lceil \log_m n \rceil}\right) \subset \mathcal{O}\left(r^{\log_m n}\right)$. It should be noted that this does not always work, that is, $\mathcal{O}(f(g(n)) \neq \mathcal{O}(f(\lceil g(n) \rceil)))$ for general increasing functions f, g, with one example being an f increasing at a similar rate as that of the n-factorial function, that may result in

$$f(\lceil g(n) \rceil) = f(g(n) + a) = (g(n) + a) \cdot f(g(n)) \notin \mathcal{O}(f(g(n)))$$

for $a < 1$.

But from this, the time complexity corresponding to $\langle n,n,n\rangle$ is $\mathcal{O}\left(r^{\log_m n}\right) = \mathcal{O}\left(n^{\log_m r}\right)$, where in particular Strassen's algorithm results in $\mathcal{O}\left(n^{\log_2 7}\right)$, since $r \leq 7$ in the case $n = 2$. More importantly to this case, however, is that

$$\tilde{\omega} \leq \log_m r, \text{ for any } m > 1. \tag{11}$$

Finally, an m resulting in a small r relative to m^{ω} is chosen, done in order to compare $\tilde{\omega}$ and ω. Given that ω is an infimum, it holds, for some α and for each $\varepsilon > 0$, that $r = R(\langle m,m,m\rangle) \leq \alpha m^{\omega+\varepsilon}$ for all large m. Then $\log_m r \leq \omega + \varepsilon + \log_m \alpha$, but it should be noted that the last term can be made arbitrarily small given the choice of a large enough m since α is constant. This means that for each $\varepsilon > 0$, $\tilde{\omega} \leq \omega + \varepsilon$, and so $\tilde{\omega} \leq \omega$, resulting in the following.

Proposition 3. *The exponent of matrix multiplication describes the time complexity of matrix multiplication according to* $\omega = \tilde{\omega}$.

All this gives rise to two main tracks for the development of faster matrix multiplication that are, although different, very much overlapping. The first track is having the goal of simply minimizing ω, which corresponds to figuring out how fast multiplication can be in theory. The second track is finding new decomposition of individual tensors $\langle l,m,n\rangle$ giving new bounds on their rank, which results in a new multiplication algorithm. The second track of course gives bounds on the time complexity, and given the nature of ω, bounding it non-trivially must involve manipulations of matrix multiplication tensors, and as such the first track also creates new algorithms, at the very least implicitly. The difference of the methods lies more in their respective goals and practicality. Decompositions of tensors will result in algorithms that are often both simple and practical, with the common trait to simply do fewer multiplications at the cost of more additions, whereas algorithms related to bounds of ω will in general not care about practicality. There may be, and throughout the history of algebraic complexity theory very much have been, multiplication algorithms that are unstable, that is, error in inputs are propagated, or involve constants so large that they are only optimal for impractically large matrices, even though they give a low bound on the exponent[9].

As seen in the previous bounding in (11), knowing the rank of a specific matrix multiplication tensor $\langle N,N,N\rangle$, or being able to bound it from above by r, results in a bounding of

$$\omega \leq \log_N r. \tag{12}$$

However, this also works for all $\langle l,m,n\rangle$ which is due to the fact that a $\langle N,N,N\rangle$ tensor can be achieved using manipulations on a $\langle l,m,n\rangle$ tensor. From the definition of rank, the value is independent of the order of the x,y,z variables. In other words, tensors $t \in X\otimes Y\otimes Z$ have the same rank as the corresponding tensors of $X\otimes Z\otimes Y$ or any other permutation of $\{X,Y,Z\}$, where the factor of each rank one term has been permuted appropriately. As such, $R(\langle l,m,n\rangle) = R(\langle \pi(l,m,n)\rangle)$ for any permutation π of three elements. This is because $\langle \pi(l,m,n)\rangle$ will either correspond to a permutation of $M_{l\times m}, M_{m\times n}, M_{n\times l}$, transposing the matrices, or both, and neither will change the rank of the tensor.

One can also define a *tensor product* $t\otimes t'$ of $t = \sum_{i,j,k} t_{ijk} x_i \otimes y_j \otimes z_k \in K^l \otimes K^m \otimes K^n$ and

$t' = \sum_{i',j',k'} t'_{i'j'k'} x'_{i'} \otimes y'_{j'} \otimes z'_{k'} \in K^{l'} \otimes K^{m'} \otimes K^{n'}$ as

$$t \otimes t' := \sum_{i,j,k,i',j',k'} t_{ijk} t'_{i'j'k'} x_{ii'} \otimes y_{jj'} \otimes z_{kk'} \in K^{ll'} \otimes K^{mm'} \otimes K^{nn'}$$

where basis vectors $x_{ii'}$ of $K^{ll'}$ correspond to $x_i \otimes x'_{i'}$ of the isomorphic $K^l \otimes K^{l'}$, and similarly for the y- and z-variables. The direct main consequence of this will be that

$$R(t \otimes t') \leq R(t)\, R(t'), \tag{13}$$

seen when multiplying t and t' under law of distribution.

Applied to matrix multiplication, one can show the relation

$$\langle a, b, c \rangle \otimes \langle a', b', c' \rangle = \langle aa', bb', cc' \rangle.$$

Remembering that matrix multiplication tensors on form

$$\langle a, b, c \rangle = \sum_{i,j,k,l,m,n} t_{ijklmn} x_{ij} \otimes y_{kl} \otimes z_{mn}$$

have coefficients $t_{ijklmn} = \delta_{in}\delta_{jl}\delta_{km}$ from (3), the tensor product $\langle a, b, c \rangle \otimes \langle a', b', c' \rangle$ will have coefficients $(\delta_{in}\delta_{i'n'})(\delta_{jl}\delta_{j'l'})(\delta_{km}\delta_{k'm'})$, which corresponds to the coefficients of $\langle aa', bb', cc' \rangle$ when using four indices per factor variable. Then, given that $R(\langle l, m, n \rangle) \leq r$, using both permutation-invariance and tensor products described above, (13) results in

$$R(\langle l, m, n \rangle \otimes \langle m, n, l \rangle \otimes \langle n, l, m \rangle) = R(\langle lmn, lmn, lmn \rangle) \leq r^3,$$

where replacing $N = lmn$ in (12) gives the following result.

Proposition 4. *If $R(\langle l, m, n \rangle) \leq r$, then $\omega \leq 3 \log_{lmn} r$.*

It should finally be noted that at first glance, ω is dependent on tensors $\langle n, n, n \rangle$ and as such dependent on the underlying field of the matrix multiplication. It is a valid concern, since, as will be seen later in the text, the rank of some tensors with coefficients being only ones and zeroes, that thus exist for every field K, will be dependent on the field. While it is beyond the scope of this text, where different ω:s will not be considered, it can be shown that fields with the same characteristic will have the same exponent[10]. Thus, for most practical purposes, the ω of interest will be a single unique one, for fields $\mathbb{R}$ and $\mathbb{C}$.

3.2 Relaxations of rank

3.2.1 Border rank

Methods for bounding ω include introducing new relaxations of tensor rank, with the most common being so-called border rank of a tensor. In essence, this is the lowest rank among all approximations of the tensor. More formally, for a tensor $t \in K^l \otimes K^m \otimes K^n$ and an integer h, define

$$R_h(t) := \min \left\{ r \ : \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \in \varepsilon^h t + \mathcal{O}\left(\varepsilon^{h+1}\right), \ u_i \in K[\varepsilon]^l \text{ and similarly for } v_i, w_i \right\} \quad (14)$$

where here $\mathcal{O}\left(\varepsilon^k\right)$ is shorthand for $\cup_s \mathcal{O}\left(\varepsilon^k s\right)$ for all tensors s. Here, ε is a formal variable, but should in practice and for algorithms be viewed as small numbers, and as such the tensor

$$\sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \in \varepsilon^h t + \mathcal{O}\left(\varepsilon^{h+1}\right)$$

here is viewed, and will be defined, as an *approximation* of degree h of t. Also, this means that

$$\mathcal{O}\left(\varepsilon^h\right) \subset \mathcal{O}\left(\varepsilon^{h'}\right) \quad \text{for } h \geq h' \quad (15)$$

which follows directly from fields of characteristic zero such as $\mathbb{R}$. For other fields, however, (15) needs to be said to hold by definition, as part of defining (14). It can also be seen by comparing tensors $s \in \varepsilon^h t + \mathcal{O}\left(\varepsilon^{h+1}\right)$ and $s \cdot \varepsilon$, that

$$R_h(t) \leq R_{h'}(t)$$

for $h \geq h'$. Then, the *border rank* of t is defined as $\underline{R}\,(t) := \min_h R_h\,(t)$. With the case $h = 0$ overlapping exactly with the definition of original rank, border rank becomes a relaxation of rank with $\underline{R}\,(t) \leq R\,(t)$.

Border rank is of course useful for applications involving approximation, but are also used in the context of faster matrix multiplication due to the fact that $R\,(t)$ can be bounded using $R_h\,(t)$, which comes from an algorithm computing t using an approximation with ε-variables. As such border rank can be used to bound ω.

If $R_h\,(t) = r$, there exists a tensor in $t' \in K[\varepsilon]^l \otimes K[\varepsilon]^m \otimes K[\varepsilon]^n$ such that

$$t' = \sum_{i'=1}^{r} u_{i'} \otimes v_{i'} \otimes w_{i'} \in \varepsilon^h t + \mathcal{O}\left(\varepsilon^{h+1}\right),$$

but here one can exclude most $u_{i'}, v_{i'}, w_{i'}$. Namely, $u_{i'}, v_{i'}, w_{i'}$ can all be assumed to be polynomials of degree h. In other words,

$$u_{i'} = \sum_{i=0}^{h} \varepsilon^i u_{i',i}$$

for $u_{i,i'} \in K^l$, and similarly for $v_{i'}$ and $w_{i'}$. This is due to the fact that the terms of t' with higher degrees would be guaranteed to fall under $\mathcal{O}\left(\varepsilon^{h+1}\right)$. But then t' can be written, using $u_{i,i'}$:s etc, as

$$t' = \sum_{i'=1}^{r} \sum_{(i,j,k)\in[0,h]^3} \varepsilon^{i+j+k} u_{i,i'} \otimes v_{j,i'} \otimes w_{k,i'}.$$

Then, when computing t exactly from t', only the terms with $i + j + k = h$ are used, of which there are $\binom{h+2}{2}$ for each i'. The number of those rank one terms gives an upper bound of the rank of t, which results in the following inequality.

Proposition 5. *For each tensor t and non-negative integer h, $R(t) \leq \binom{h+2}{2} R_h(t)$.*

There is also a border rank parallel to Proposition 1, where s is a *degeneration* of t if there is an approximation of s that is a restriction of t. In the literature, this is usually denoted $s \trianglelefteq t$,[10] and leads to $\underline{R}(s) \leq \underline{R}(t)$. It should be noted that a tensor being equal to another means that it is still an approximation, just with $h = 0$. Therefore, restrictions fall under degenerations.

3.2.2 Alternative definitions of border rank

There are alternative definitions for border rank, that are, too, grounded in approximating tensors. For fields $\mathbb{R}$ and $\mathbb{C}$, border rank can also be defined by sequences. Then, t is said to have border rank $\underline{R}(t) = r$ if r is the smallest number such that a sequence of tensors $\{t_i\}_{i=1}^{\infty}$, with $R(t_i) \leq r$ for each i, converges to t.[10]

Yet another definition is to again use converging sequences, but under the Zariski topology instead, where the field K being algebraically closed is needed. This means that t is said to have border rank $\underline{R}(t) = r$ if r is the smallest number such that t is in the Zariski closure of $\{t' : R(t) \leq r\}$. [7] As such, methods from commutative algebra and algebraic geometry can be used to bound tensor ranks as well.

These definitions each have their own strengths, where the ε-definition has a directly intuitive connection to concrete approximations of tensors, and it in theory works for any field K, even though the interpretation of ε as an "infinitely small element" cannot always be made.

3.2.3 Support rank

By fixing bases $\{x_i\}_i, \{y_j\}_j, \{z_k\}_k$ for vector spaces X, Y, Z respectively, a basis $\mathcal{B} = \{x_i \otimes y_j \otimes z_k\}_{i,j,k}$ is formed for the tensor product $X \otimes Y \otimes Z$. Unless stated, each chosen basis of a tensor product will be assumed to be of this form. Then, the *support* of a tensor $t = \sum_{i,j,k} t_{ijk} x_i \otimes y_j \otimes z_k \in X \otimes Y \otimes Z$ with respect to basis $\mathcal{B}$ is the set

$$\text{supp}_{\mathcal{B}}(t) := \{(i,j,k) \in [\dim X] \times [\dim Y] \times [\dim Z] : t_{ijk} \neq 0\} \in 2^{\mathcal{B}},$$

where then $\text{supp}_{\mathcal{B}}$ is a map $X \otimes Y \otimes Z \to 2^{\mathcal{B}}$. This defines an equivalence relation

$$t \sim_{s,\mathcal{B}} t' \text{ if and only if } \text{supp}_{\mathcal{B}}(t) = \text{supp}_{\mathcal{B}}(t')$$

with equivalence classes $[t]_{s,\mathcal{B}}$ called *support classes*. The index $\mathcal{B}$ is dropped from the notation if the basis is clear from the context. The *size* of a support or support class is then defined to be simply $|\,\text{supp}_{\mathcal{B}}(t)\,|$. It should be noted that by setting one of the vector space dimensions to one, this definition becomes the support of a matrix, with respect to a particular basis.

Then, given the common definition of rank for three-dimensional tensors, a relaxation of rank called *support rank*, first defined by Umans and Cohns in 2012[11], with regards to basis $\mathcal{B}$ can be defined as

$$R_{s,\mathcal{B}}(t) := \min\{R(t') : t' \sim_{s,\mathcal{B}} t\}.$$

Again, $\mathcal{B}$ is dropped from the notation if the basis is clear from the context. Since the support rank only depends on the support, one can also write $R_s([t]_s) = R_s(t)$, or simply $R_s(S)$, for some support $S \in 2^{\mathcal{B}}$. For a support or support class, the *possible ranks* can be defined as the set

$$R([t]_s) := \{R(t') : t' \in [t]_s\}$$

and its *maximum rank* the number $R_m([t]_s) := \max R([t]_s)$.

Similarly to the other relaxation border rank, by definition it holds for support rank that $R_s(t) \leq R(t)$ for all tensors t. However, it will be shown later that support rank and border rank cannot be compared in the same way.

At first glance the support rank seems much harder to compute than regular rank due to the fact that for support rank, the rank of every tensor in a support class need to be known. In some cases, there exist shortcuts, such as with the equivalence

$$R_s(S) = 1 \text{ if and only if } R(1(S)) = 1$$

where the *one-tensor* of support S, denoted $1(S)$, is defined to be the tensor with support S having only ones and zeroes as coefficients. The if-part holds due to the minimum rank of a non-zero tensor is one, and the zero tensor is the only element in its support class, and so $1 \leq R_s(S) \leq R(1(S)) = 1$. For the only if-part, suppose t' has rank one and support S. By definition of rank one, there are scalars $a'_i, b'_j, c'_k \in K$ such that $t' = \sum_{i,j,k} a'_i b'_j c'_k x_i \otimes y_j \otimes z_k$ in the chosen basis, and noting that making a change of variables

$$a'_i \mapsto a_i = \begin{cases} 1\ , & a'_i \neq 0, \\ 0\ , & a'_i = 0, \end{cases}$$

and similarly for b_j, c_k, leads to transforming t' to $1(S)$. But then there are scalars $a_i, b_j, c_k \in K$ such that $1(S) = \sum_{i,j,k} a_i b_j c_k x_i \otimes y_j \otimes z_k$ and so $R(1(S)) = 1$.

Another way to phrase this is that the one-tensor of a support class has the minimal rank if the support rank is one, and so it is enough to check the one-tensor when verifying a support rank being one. The same does not hold for higher support ranks however. Using analogue definitions for support rank on two-dimensional tensors, by setting one of the dimensions of the spaces X, Y, Z to one, the matrix

$$A(a) := \begin{pmatrix} 0 & 1 & 1 \\ a & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

has full matrix rank three if $a = 1$. This is because its determinant is

$$\det A(a) = -a \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = a + 1$$

and thus non-zero. However, setting $a = -1$ leads to a zero determinant and a rank strictly smaller than three, and so the one-tensor of the class, $A(1)$, does not have the minimal rank of the class.

While both tensor rank and support rank definitions are dependent on choices of basis, only common rank is basis-invariant, while it will not be for support rank. Given a tensor t, there will always be changes of bases $\{x_i\} \to \{x_i'\}$ in X and similarly for Y and Z, such that t in this new basis $\mathcal{B}'$ has no zero terms and a full support. The one-tensor of a full support has rank one, and so $R_{s,\mathcal{B}'}(t) = 1$. If support rank was basis-invariant, then every tensor should have support rank equaling one in every basis, but this is not the case. For instance, a matrix multiplication tensor written on the usual form $\langle n,n,n\rangle = \sum_{i,j,k} x_{ij} \otimes y_{jk} \otimes z_{ki}$ is the one-tensor of its support, but as seen before this rank has been bounded from below by n^2. Therefore, if $n > 1$, $\langle n,n,n\rangle$ cannot have support rank equaling one in some basis, thus making support rank not basis-invariant.

It should be noted that while support rank has the problem of being basis dependent, it can still find, and has found, use within computational mathematics and algebraic complexity theory in particular, since a specified explicit basis is needed for any computation or problem related to computation. Furthermore, one can define ω_s analogous to ω by just switching out rank notation $R(\langle n,n,n\rangle)$ to $R_s(\langle n,n,n\rangle)$, and it can be shown that they are related by

$$\omega \leq (3\omega_s - 2)/2.$$

As such, support rank can be used for upper bounds of ω even though it is a relaxation of rank,[11] a similar story as to that of border rank.

More recently, the support rank of the matrix multiplication tensor $\langle 2,2,2\rangle$ has been computed to be exactly seven, and same for a so-called *border support rank* which is the minimum border rank within the support class.[7]

3.3 Recent developments on matrix multiplication

In 2022, new upper bounds for ranks of numerous matrix multiplication were found using machine learning.[12] The actual resulting algorithms from the ranks - not the methods used to find them - are simple and in the vein of Strassen's first algorithm in 1969, that is, at a cost of making more additions making fewer multiplications. So while not bounding ω better than before, new practical algorithms have been discovered.

On the other track, with the goal of just bounding ω, further records have been set, being ≈ 2.373 as of 2020 and popularly hypothesised to be as low as 2.[9] The most interesting development, however, is actually a setback. The main method of bounding ω, dubbed *the universal method*, has been found to have limits, even though it is extremely general as a method. In essence, the universal method is from a low rank tensor t find the best degeneration mapping to a matrix multiplication-related tensor s, and then using the relation $\underline{R}(s) \leq \underline{R}(t)$ to bound ω. The lowest bounding that can be found starting from s is denoted $\omega_u(s)$, and in 2020 Alman showed that there existed degeneration methods that have been used historically and that are complete with respect to $\omega_u(s)$. In this case, complete means that if $\omega_u(s) = 2$, then using one of these degeneration methods on s, if it could be applied to it, would yield a bound of $\omega \leq 2$. Thus, the fact that a tensor s did not yield a bound of $\omega \leq 2$, means that there is no way of degenerating s using the universal method and achieving $\omega_u(s) = 2$. What is the most troubling is that all new records set on ω since 1990 have used a family called Coppersmith-Winograd tensors

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

where q is any positive integer and σ a permutation on $[q]$. These are applicable to and have been used by the so-called Laser method, which is complete.[9] This leads to two paths, with the first being finding some other method not falling under the umbrella of the universal method. This seems the most unlikely however, given that the universal method is so general, and it is difficult to come up with a way of bounding ranks without using linear mappings, which is what degeneration methods amount. The other and more likely choice is to find new families of low ranking tensors that can degenerate into matrix multiplication tensors, although as stated, one family has dominated the field for more than 30 years.

4 Rank structures of support classes

With non-zero support classes of $X \otimes Y \otimes Z$ in basis $\mathcal{B}$, there are the two extreme examples of the support S being the empty set, and the support S' being all basis elements in $\mathcal{B}$, corresponding to having support size either one or maximum. In both cases, the support rank is one, but for the full support S', its possible ranks are, as seen earlier, all possible ranks in the tensor product space. At the same time, there exists a symmetry among support classes with regards to support size. This is because choosing which coefficients should be non-zero for a support, is the same as choosing which coefficients should be zero. There is then a bijection between support classes of size k and support classes of size $\dim(X \otimes Y \otimes Z) - k$, by just switching non-zero coefficients to zero, and vice versa.

This leads to the question what the possible rank structures of support classes are between the two extreme supports S and S', and if they are somehow dependent on the support size of the classes. Furthermore, will then the symmetry among support classes with regards to support size be apparent in the results?

To answer these questions fully, a classification of possible ranks of support classes in a given tensor product space must be made. The number of different support classes in a tensor product $K^l \otimes K^m \otimes K^n$ of size k will be $\binom{lmn}{k}$, since the classes are defined by selecting which k of the lmn basis tensors that should have a non-zero coefficient. This number $\binom{lmn}{k}$ will generally be large, where even for the smallest dimension eight that gives a non-degenerate three-dimensional tensor product structure, there are $\binom{8}{4} = 70$ support classes of size four.

4.1 Finding equivalent supports

However, for a given support size, there will be different classes that are equivalent using permutations of the variables between the three vector spaces. For a permutation isomorphism, which will be defined more formally below, the elements of a support class will all be mapped to the same new support class, since different non-zero coefficients of an element will not affect the support of the image when using only permutation. For example, tensors of the form

$$t_1 = \alpha x_1 y_1 z_1 + \beta x_2 y_3 z_3 + \gamma x_1 y_3 z_4 + \delta x_2 y_2 z_2 + \varepsilon x_1 y_2 z_3 \in K^2 \otimes K^3 \otimes K^4,$$
$$t_2 = \alpha x_1 y_2 z_2 + \beta x_3 y_1 z_1 + \gamma x_4 y_1 z_2 + \delta x_2 y_3 z_1 + \varepsilon x_3 y_3 z_2 \in K^4 \otimes K^3 \otimes K^2$$

are equivalent, using permutations $x_1 \to x_2$ and $y_1 \to y_2 \to y_3 \to y_1$ and then switch all x with z and vice versa, where the first mapping is permuting variables within vector spaces, and the second is permuting the vector spaces $K^2 \to K^4$. Since t_1 and t_2 are equivalent, their respective support classes will have the same rank structure for their elements. In other words, how the non-zero coefficients of an element in a class affect the rank of the element will be exactly the same. This means that the number of different rank structures of support classes of size k will generally be less or equal to $\binom{lmn}{k}$.

What will from here on be referred to as *permutation isomorphisms* are linear maps constructed as follows. First, by choosing three permutations $\sigma_x, \sigma_y, \sigma_z \in S_n$ acting on the indices of x:s, y:s, and z:s respectively, which creates a mapping $x_i \mapsto x_{\sigma_x(i)}$ and so on. Then, by choosing a permutation $\pi \in S_3$ for the vector spaces, which permutes the symbols $\{x, y, z\}$ and the corresponding vector spaces $\{X, Y, Z\}$, creating a mapping $x_i \mapsto \pi(x)_i$ from X to $\pi(X)$ and so on. These four mappings act on basis elements of X, Y, Z and can thus be made to be linear. The final *permutation isomorphism*

$$\begin{aligned} \varphi : X \otimes Y \otimes Z &\longrightarrow \pi(X) \otimes \pi(Y) \otimes \pi(Z), \\ x_i y_j z_k &\longmapsto \pi(x)_{\sigma_x(i)} \otimes \pi(y)_{\sigma_y(j)} \otimes \pi(z)_{\sigma_z(k)} \end{aligned} \tag{16}$$

is then defined as the composition of these four linear maps corresponding to $\sigma_x, \sigma_y, \sigma_z, \pi$, where the composition order is first applying $\sigma_x, \sigma_y, \sigma_z$, and then π.

It should be noted that when $X \cong Y \cong Z$, the composition of $\sigma_x, \sigma_y, \sigma_z, \pi$ could a priori be made the other way around, but this would result in just another φ composed as per definition above, but with index permutation $\sigma_{\pi(x)}$ instead of σ_x and so on. This means that the set of permutation isomorphisms, or in this case permutation *auto*morphisms, on tensor spaces $K^n \otimes K^n \otimes K^n$ defines a group G_n.

From the definition of permutation automorphisms, the group size $|G_n|$ is at most $3!\,(n!)^3$, and will be exactly that if the choices for the permutations are independent of each other. Assume, for a contradiction that there are $(\sigma_x, \sigma_y, \sigma_z, \pi) \neq (\sigma'_x, \sigma'_y, \sigma'_z, \pi')$ such that their respective automorphisms φ and φ' are equal. If $\pi = \pi'$, then from (16) and since π is injective, it holds that $(\sigma_x(i), \sigma_y(j), \sigma_z(k)) = (\sigma'_x(i), \sigma'_y(j), \sigma'_z(k))$ for all i, j, k, which is a contradiction. If $\pi \neq \pi'$ where, say, $\pi(x) \neq \pi'(x)$, then since $\varphi = \varphi'$ it most hold that

$$\sigma_{\pi^{-1}(x)}(i) = \sigma'_{\pi'^{-1}(x)}(j) \text{ for all } i, j = 1, \ldots, n.$$

But for $n > 1$, this results in $\sigma_{\pi'^{-1}(x)}$ not being a well-defined function since $\sigma_{\pi'^{-1}}(1)$ takes all n different values $\sigma'_{\pi'^{-1}(x)}(j)$. This finally means that $|G_n| = 3!\,(n!)^3$, and it should be noted that, as the dimension n increases, the group size will increase much faster than the number of support classes of a given size $|X_{n,k}| = \binom{n^3}{k}$, since the latter is a polynomial in variable n if k is fixed.

While there is a clear bijection between G_n and $S_n^3 \times S_3$, this is not a group isomorphism. The latter group has a normal subgroup $A_n^3 \times A_3$ made of alternating groups, which leads to a quotient group $\left(S_n^3 \times S_3\right) / \left(A_n^3 \times A_3\right) \cong (\mathbb{Z}/2\mathbb{Z})^4$ since $|A_n| = |S_n|/2$. The quotient group needs at least four generators and as such $S_n^3 \times S_3$ cannot be generated by fewer than four elements. However, it will be seen later that G_2 can be generated using only three elements. If the bijection between G_n and $S_n^3 \times S_3$ was a group isomorphism, then the subgroup $G_2 \subset G_n$ would be isomorphic to $S_2^3 \times S_3 \subset S_n^3 \times S_3$, which is not possible.

Finally, using permutation automorphisms, the number of different rank structures may be computed using Burnside's lemma for tensor spaces $K^n \otimes K^n \otimes K^n$. Focusing on tensor spaces $K^n \otimes K^n \otimes K^n$, the set of permutation automorphisms will form a group

$$G_n = \{g_\varphi : \varphi \text{ is a permutation automorphism on } K^n \otimes K^n \otimes K^n\}.$$

With the set $X_{n,k} = \{\text{support classes of } K^n \otimes K^n \otimes K^n \text{ of size } k\}$, there is then a group action of G_n acting on $X_{n,k}$ defined by

$$\begin{aligned} G \times X_{n,k} &\longrightarrow X_{n,k}, \\ (\varphi, x) &\longmapsto \varphi * x = \varphi(x). \end{aligned}$$

Then an orbit in $X_{n,k}/G_n$ of this action will be a set of support classes with the same rank structure. For such an orbit T, which will also be called a *rank structure*, the notation $R_s(T)$, $R_m(T)$, $R(T)$, for support rank, maximum rank, and possible ranks respectively, is defined similarly as for support classes before.

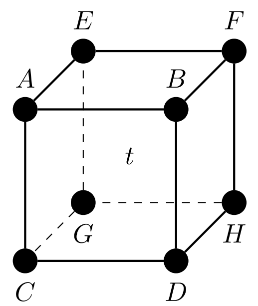
There will be some shortcuts when trying to find all orbits. Firstly, the sets $X_{n,0}/G_n$ and $X_{n,1}/G_n$ are singletons since both $X_{n,0}$ and $X_{n,1}$ are singletons. Secondly, there is due to symmetry a bijection between orbits of $X_{n,k}/G_n$ and $X_{n,n^3-k}/G_n$ by switching between non-zero-coefficients and zero-coefficients in a tensor. This does not however mean that two corresponding orbits of $X_{n,k}/G_n$ and $X_{n,n^3-k}/G_n$ will be equivalent and have the same rank structure.

4.2 Support orbits in $K^2 \otimes K^2 \otimes K^2$

In the special case $n = 2$, tensors can be visualised as cubes where coefficients of base terms $x_i y_j z_k$ correspond to vertices, where for example

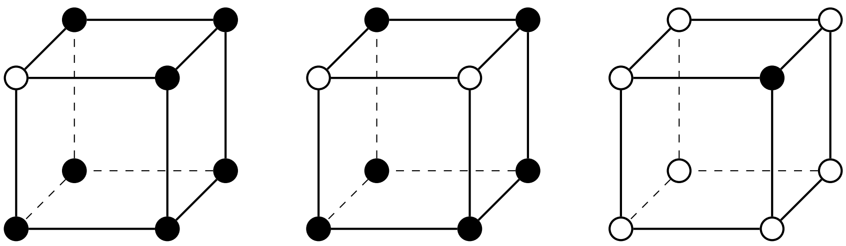
$$t = Ax_1y_1z_1 + Bx_1y_2z_1 + Cx_2y_1z_1 + Dx_2y_2z_1 + Ex_1y_1z_2 + Fx_1y_2z_2 + Gx_2y_1z_2 + Hx_2y_2z_2$$

is visualized as



From now on, white vertices will correspond to a coefficient being zero, and black vertices will, unless otherwise stated, correspond to non-zero coefficients. Furthermore, vertex labels as seen above will be omitted either if the vertex is white, or if its non-zero coefficient value is irrelevant or arbitrary, and the same letters will usually label the same vertex. For example, A is the coefficient for $x_1y_1z_1$ and C the coefficient of $x_2y_1z_1$ even if the tensor has a white vertex at $x_1y_2z_1$, where B would be.

With this, support classes can be visualised as cubes, such as the following three classes.

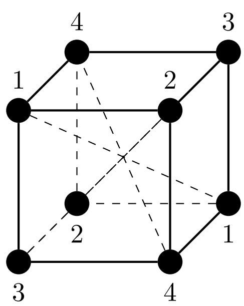


From this perspective, the permutation automorphisms of G_2 preserves distance between any two different vertices $x_iy_jz_k$ and $x_{i'}y_{j'}z_{k'}$. The vertices can only differ by one, two, or three indices, corresponding to three types of distance and, in all cases, a permutation automorphism will not change the difference in indices, since it can only switch indices $1 \to 2$ or permute the x, y, z-names. This means that G_2 is a subset of the automorphism group on the cube, but G_2 also includes that group's three generators: rotation in the x-axis, rotation in the y-axis, and reflection. As such G_2 is actually isomorphic to the automorphism group on the cube.

Onto computing the orbits of $X_{2,k}/G_2$, Burnside's lemma

$$|X_{2,k}/G_2| = \frac{1}{|G_2|} \sum_{g \in G_2} |X_{2,k}^g| \tag{17}$$

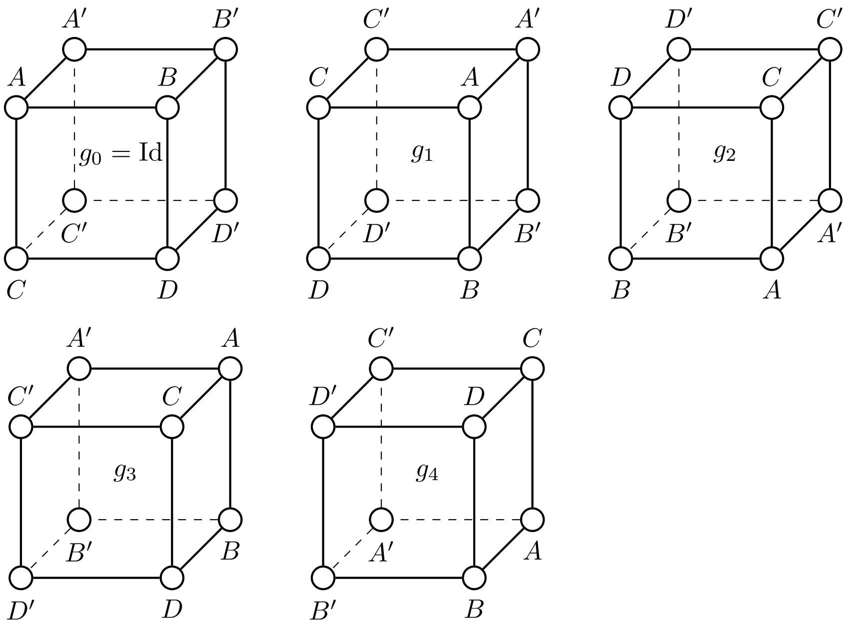
is applied, where the fix-points of a group element g is $X_{2,k}^g = \{x \in X_{2,k} : g * x = x\}$. Here, a useful method is to partition the group into its conjugacy classes, as two conjugate elements have the same fix-points due to $\left(h^{-1}gh\right) * x = x$ if $g * x = x$. The group of automorphisms on the cube is isomorphic to $S_4 \times S_2$, where the subgroup of rotations is in turn isomorphic to $S_4 \times \{\mathrm{Id}\} \cong S_4$, where rotations correspond to permutations of longest diagonals between vertices of the cube[13], visualized as



(18)

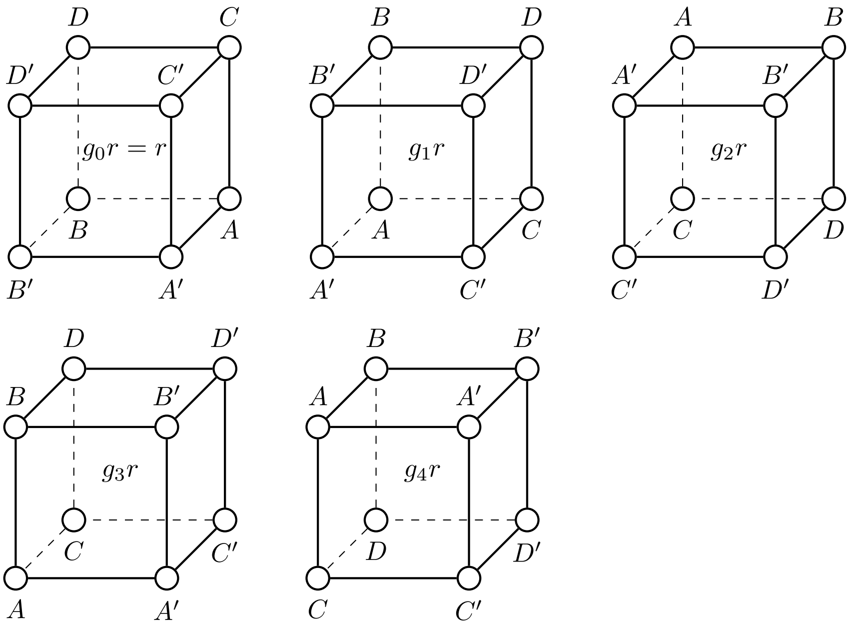
The element $(\mathrm{Id}, (12)) \in S_4 \times S_2$ corresponds to reflecting across all four longest diagonals, i.e, switching i with the other i in (18).

The conjugacy classes of G_2 are then those that correspond to the conjugacy classes of S_4, of which there are five,[13] plus a copy of each such class C which corresponds to $C \times \{(12)\} \subset S_4 \times S_2$, totaling in ten conjugacy classes. The five rotation conjugacy classes are as follows: the identity element $g_0 = \mathrm{Id}$; the class $[g_1]$ where its representative g_1 is a rotation of $\pi/2$ around an axis perpendicular to a cube side, where $|[g_1]| = 6$ as each such axis has two such rotations; the class $[g_2]$ where $g_2 = g_1^2$, that is, a rotation of π around an axis perpendicular to a cube side, where then $|[g_2]| = 3$; the class $[g_3]$ where g_3 is a rotation of $2\pi/3$ around an axis that follow the longest diagonals between edges of the cube, where each of the four diagonals has two such rotations, and so $|[g_3]| = 8$; and finally $[g_4]$ where g_4 is a rotation of π around an axis that follow the line between the middle points of two furthest away edges of the cube, where then $|[g_4]| = 6$. The elements g_i can be visualized as the result when g_i has acted on a cube, with vertices named $A, B, C, D, A', B', C', D'$, in the following figures.



(19)

The remaining conjugacy classes of G_2 can be described by representatives $g_i \cdot r$ with $|[g_i]| = |[g_i \cdot r]|$, where r is the reflection corresponding to $(\mathrm{Id}, (12))$ described above. This results in representatives of classes $[g_i \cdot r]$ visualized in the same way as in (19) in the following figures.



(20)

Now the size of each set of fix-points $X^g_{2,k}$ can be computed for support sizes $k = 2, 3, 4$. This is done by seeing how many ways one can color k vertices in each cube in (19) and (20), where the label of the colored vertex stays the same. The results can be seen in Table 1 below. The cases $k = 0, 1$ is not included since the number of orbits $|X_{2,k}/G_2| = 1$ is then already known.

g	$\lvert[g]\rvert$	$\lvert X^g_{2,2}\rvert$	$\lvert X^g_{2,3}\rvert$	$\lvert X^g_{2,4}\rvert$
g_0	1	$\binom{8}{2}$	$\binom{8}{3}$	$\binom{8}{4}$
g_1	6	0	0	2
g_2	3	4	0	6
g_3	8	1	2	4
g_4	6	4	0	6
g_0r	1	4	0	6
g_1r	6	0	0	2
g_2r	3	4	0	6
g_3r	8	1	0	0
g_4r	6	8	12	14

Table 1: Number of fix-points for conjugacy class representatives g_i and $g_i r$ for $i = 0, \dots, 4$

Then the number of orbits can be computed using Burnside's lemma (17), and using conjugacy classes the formula becomes

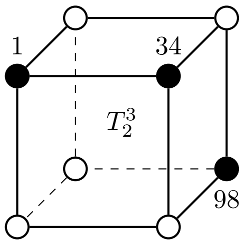
$$|X_{2,k}/G_2| = \frac{1}{|G_2|}\left(\sum_{i=0}^{4} |[g_i]| \cdot \left(|X^{g_i}_{2,k}| + |X^{g_i r}_{2,k}|\right)\right).$$

Using values in Table 1, as well as the group size $|G_2| = 3! \cdot (2!)^3 = 48$, the number of orbits can be calculated to be

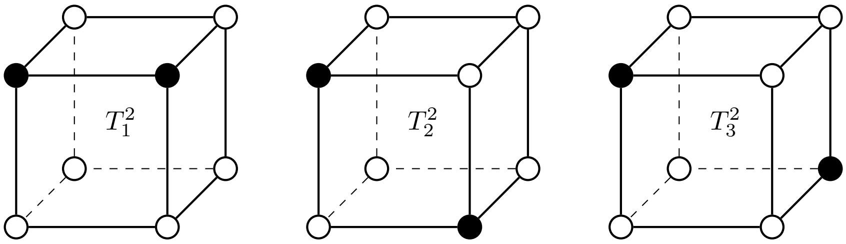
$$|X_{2,2}/G_2| = 3\ ,\ \ |X_{2,3}/G_2| = 3\ ,|X_{2,4}/G_2| = 6,$$

which are all much less than $|X_{2,k}| = \binom{8}{k}$ for $k = 2, 3, 4$.

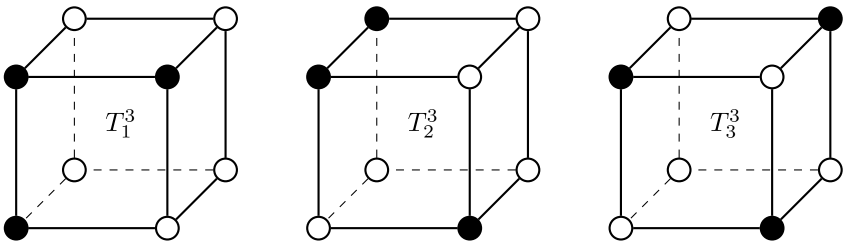
The orbits of support size k will be denoted T_i^k for $i = 1, 2, \ldots, |X_{2,k}/G_2|$, and the cubic visualisation will be that of a single support class within the orbit. Sometimes, to clarify that one tensor belongs to a specific orbit, the cube representation of the tensor will have T_i^k written inside, even if it is just one particular tensor. One example is the following single tensor.



The three orbits of support size two and the three of support size three are

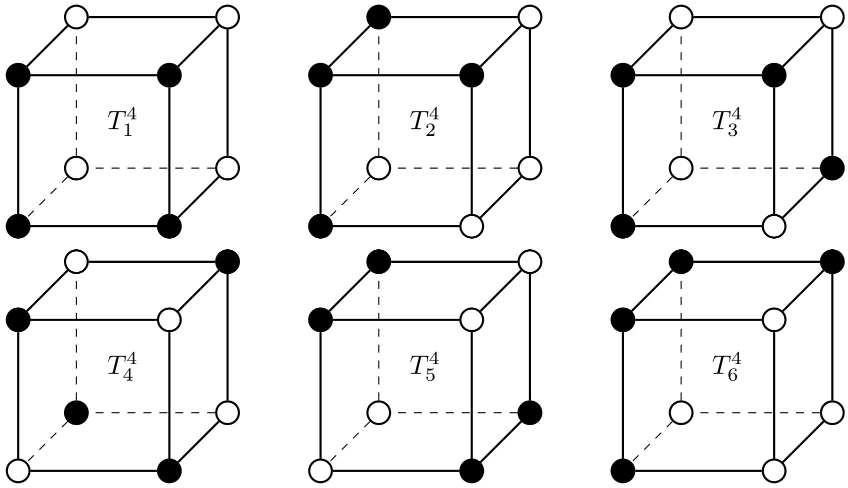


(21)



(22)

and the six of size four are the following.

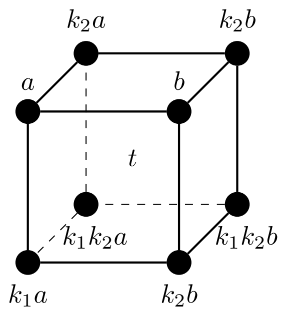


(23)

As mentioned before, for support sizes $k > 4$, the same Burnside method can be used as on sizes $8-k$, and just switching between the definition of black and white vertices, getting analogous results. An orbit T_i^k will then denote T_i^{8-k} but with inverted vertex colors, for $k > 4$.

4.3 Classification of rank structures of supports in $K^2 \otimes K^2 \otimes K^2$

Before classifying the rank structure of each orbit of G_2, one can make a classification of all tensors of rank one in $K^2 \otimes K^2 \otimes K^2$, that are then on the form $t = \sum_{i,j,k} a_i b_j c_k x_i y_j z_k$. From the definition in (4), it can be assumed that both $a_1 = 1$ and $c_1 = 1$, since otherwise a_1 and c_1 can be factored out of their vectors $\sum_i a_i x_i$ and $\sum_k c_k z_k$ and multiplied into $\sum_j b_j y_j$. Then, by denoting $b_1 = a,\ b_2 = b,\ a_2 = k_1,\ c_2 = k_2$, a rank one tensor, or zero tensor, in $K^2 \otimes K^2 \otimes K^2$ has a cubic visualisation



(24)

where edges may or may not be zero even though they are colored black here. It is then easy to check in particular if a tensor without full support has rank one, since then one of a, b, k_1, k_2 is zero which means an entire side of the cube, or *slice* of the tensor, is zero. A direct consequence of this is that there are no rank one-tensors among rank structures $X_{2,k}/G_2$ of support sizes $k = 5, 6, 7$.

4.3.1 Support sizes $k = 2, 3$

Now, computing rank structures for orbits starting with support size two, elements in T_1^2 and T_2^2 have two-dimensional tensor structures, as seen in (21). They will thus have the same rank as a corresponding matrix rank. For T_1^2, it is one for all elements, and for all elements in T_2^2 the rank is two. For T_3^2, there is no slice being zero, and so cannot have rank equaling one, and so every element in T_3^2 must have rank equaling two.

For support size three, T_1^3 has, by same argument as T_2^2, only ranks being two. This is because, as seen in (22), T_1^3 has a two-dimensional tensor structure. For T_2^3, it cannot have rank being one since there is no zero slice, and every element can be written as a sum of rank one tensors in T_1^2 and T_1^1, and so tensors in T_2^3 always have rank equaling two. Finally, for T_3^3, the case where all non-zero coefficients were ones was computed in Section 2.4, and for $x_1y_1z_1 + x_2y_2z_1 + x_1y_2z_2$, one can reach every element in the orbit by scaling variables according to

$$\begin{aligned}
&x_1 \mapsto Ax_1,\ y_1 \mapsto y_1,\ z_1 \mapsto z_1,\\
&x_2 \mapsto Dx_2,\ y_2 \mapsto y_2,\\
&z_2 \mapsto F/Az_2
\end{aligned}$$

for any non-zero A, D, F, and so the orbit has only one possible rank, namely that of three.

So far, the rank structures are as follows.

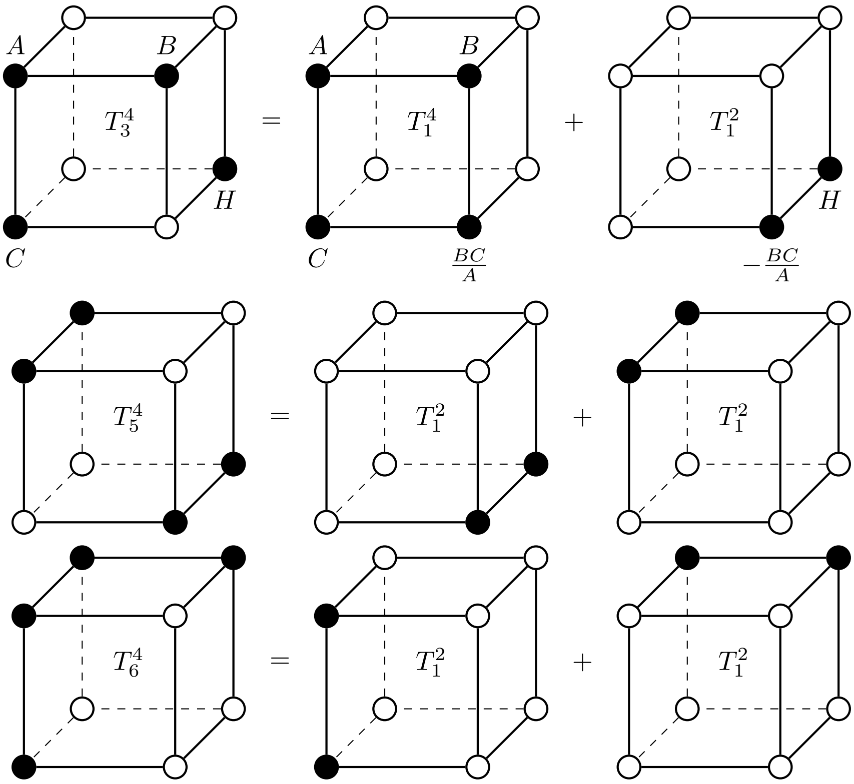
k	$\lvert X_{2,k}/G_2\rvert$	Support ranks $R_s\left(T_i^k\right)$	Max ranks $R_m\left(T_i^k\right)$
0	1	1	1
1	1	1	1
2	3	$R_s\left(T_1^2\right) = 1,\ R_s\left(T_2^2\right) = 2$ $R_s\left(T_3^2\right) = 2$	$R_m\left(T_1^2\right) = 1,\ R_m\left(T_2^2\right) = 2$ $R_m\left(T_3^2\right) = 2$
3	3	$R_s\left(T_1^3\right) = 2,\ R_s\left(T_2^3\right) = 2$ $R_s\left(T_3^3\right) = 3$	$R_m\left(T_1^3\right) = 2,\ R_m\left(T_2^3\right) = 2$ $R_m\left(T_3^3\right) = 3$

Table 2: Ranks for support orbits in $X_{2,k}/G_2$ with support sizes k

4.3.2 Support size $k = 4$

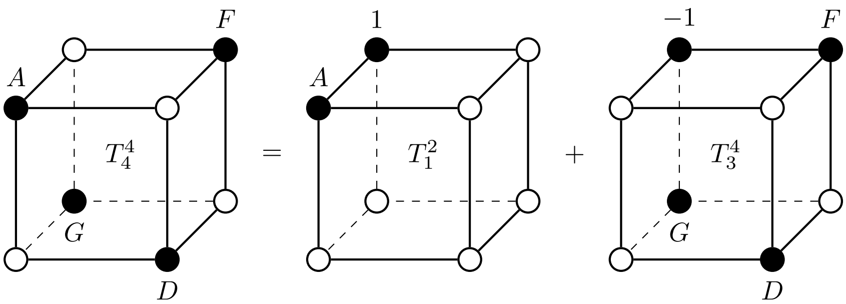
Moving on to support size four, T_1^4 is the first orbit which contain multiple ranks. Similarly to T_1^3, it has the same rank as the matrix rank of the z_1-slice, which is either one or two. Thus elements in T_1^4 all have support rank equaling one, while the maximum rank within the orbit is two, except for over the field $K = \mathbb{Z}/2\mathbb{Z}$. For the remaining five orbits, there is none with support rank one, since there is no zero slice. Assuming that a tensor $t = Ax_1y_1z_1 + Bx_1y_2z_1 + Cx_2y_1z_1 + Ex_1y_1z_2$ in T_2^4 has rank equaling two, a substitution method can be applied as to T_3^3, where t is viewed as a multi-variable polynomial in x, y, z. If t were to have rank two, then one of the two terms must depend on both z_1 and z_2, and so that term can become zero given a substitution $z_2 = \alpha z_1$ for some non-zero α. Then $t\left(z_2 = \alpha z_1\right) = \left(\left(A + E\alpha\right)x_1y_1 + Bx_2y_1 + Cx_1y_2\right)z_1$ is a rank one tensor. Since it depends on both x_1 and x_2, it can be made to zero by substitution $x_2 = \beta x_1$ for some non-zero β. Then $0 = t\left(z_2 = \alpha z_1, x_2 = \beta x_1\right) = x_1\left(\left(A + E\alpha + B\beta\right)y_1 + Cy_2\right)z_1$, but since $C \neq 0$ this is a function dependent on y_2 and thus cannot be always zero. This leads to a contradiction, and so the only rank of T_2^4 will be three.

The three orbits T_3^4, T_5^4, T_6^4 all have ranks only equaling two. This can be seen by adding rank one-elements of orbits of smaller support size in the following ways.



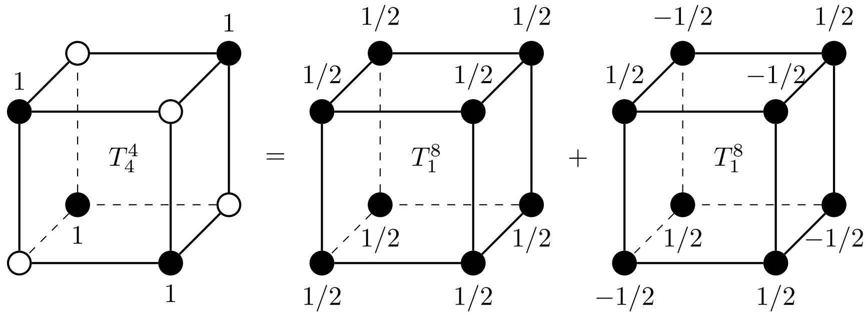
(25)

The outlier among the rank structures of support size four in (23) is the orbit T_4^4. Firstly, the maximum rank is not larger than three since a tensor in the orbit can always be written as a rank one- and rank two-tensor according to



For the one-tensor in T_4^4 and with a field K with char $K \neq 2$, the rank is two, resulting in $R_s\left(T_4^4\right) = 2$

since it can be written as



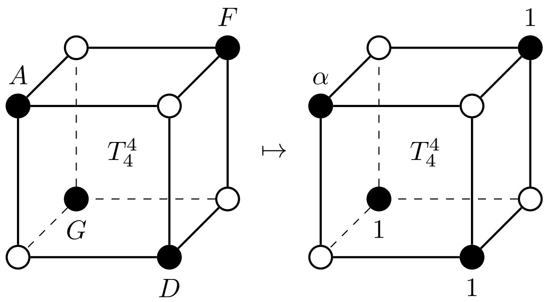
(26)

where the elements with full support have rank one since they are on the form (24).

But similarly to before, any element of T_4^4 can be transformed to another in the orbit by scaling the x, y, z-variables, but in this case, the scaled element will still have a parameter dependent on the original element. More precisely, given an element with coefficients A, D, F, G, there is a bijection

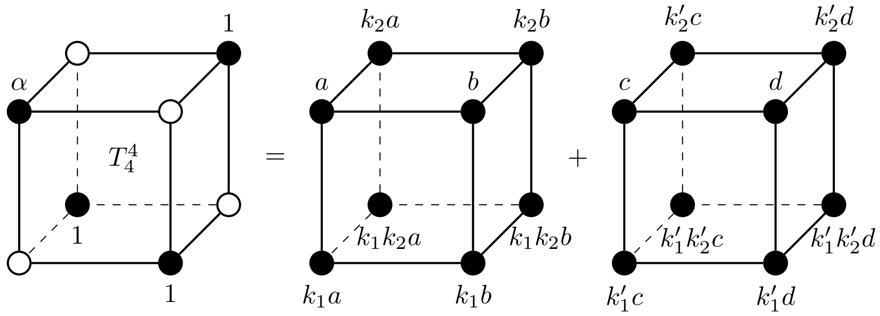
$$Gx_2 \mapsto x_2,\ y_1 \mapsto y_1,\ z_2 \mapsto z_2,$$
$$Fy_2 \mapsto y_2,\ x_1 \mapsto x_1,$$
$$\frac{D}{FG} z_1 \mapsto z_1$$

such that



(27)

where $\alpha = \frac{AFG}{D}$. For general elements of T_4^4, that is, for general α, if they have rank two, they can be written as



(28)

which corresponds to a system of eight equations in eight variables $a, b, c, d, k_1, k_2, k_1'k_2'$. Immediately, it can be seen that $d = -b$ and $c = \alpha - a$, which results in

$$(k_i - k_i')\, b = 1 \tag{29}$$

$$(k_i - k_i')\, a + k_i' \alpha = 0 \tag{30}$$

$$(k_1 k_2 - k_1' k_2')\, b = 0 \tag{31}$$

$$(k_1 k_2 - k_1' k_2')\, a + k_1' k_2' \alpha = 1 \tag{32}$$

for $i = 1, 2$. Then, noting that $b \neq 0$, inserting (29) into (30) results in $k_1' = k_2'$, which in turn results in $k_1 = k_2$ when using $k_1' = k_2'$ in (29). From (31), this means that $k_1 = \pm k_1'$ where only $k_1 = -k_1'$ can hold since they cannot cancel each other out because of (29). From this, a, b can be written as $b = 1/(2k_1)$, $a = \alpha/2$ and finally

$$k_1^2 \alpha = 1. \tag{33}$$

This means that, depending on the choice of field K and α, (28) might not have solutions. For example in $K = \mathbb{R}$, a tensor in T_4^4 has rank two if and only if $\alpha > 0$, which corresponds, according to the bijection (27), to the coefficients of a tensor in T_4^4 having all the same sign, or exactly two having the same sign. In another example, for fields with $\operatorname{char} K = 2$ only, the decomposition (26) does not exist, and there will never be two different solutions to (33) since $1 = -1$ in K. Then, no tensor in T_4^4 has rank two. In contrast, in an algebraically closed field such as $\mathbb{C}$, there is always a unique solution to (28) up to sign of k_1, and so every rank in T_4^4 is two in this case.

The rank structures can now be summarized so far in Table 3,where an exception is $K = \mathbb{Z}/2\mathbb{Z}$ where $R_s(T_4^4) = R_m(T_4^4) = 3$.

k	$\lvert X_{2,k}/G_2 \rvert$	Support ranks $R_s(T_i^k)$	Max ranks $R_m(T_i^k)$
0	1	1	1
1	1	1	1
2	3	$R_s(T_1^2) = 1,\ R_s(T_2^2) = 2$ $R_s(T_3^2) = 2$	$R_m(T_1^2) = 1,\ R_m(T_2^2) = 2$ $R_m(T_3^2) = 2$
3	3	$R_s(T_1^3) = 2,\ R_s(T_2^3) = 2$ $R_s(T_3^3) = 3$	$R_m(T_1^3) = 2,\ R_m(T_2^3) = 2$ $R_m(T_3^3) = 3$
4	6	$R_s(T_1^4) = 1,\ R_s(T_2^4) = 3$ $R_s(T_3^4) = 2,\ R_s(T_4^4) = r_{s,4}(K)$ $R_s(T_5^4) = 2,\ R_s(T_6^4) = 2$	$R_m(T_1^4) = 2,\ R_m(T_2^4) = 3$ $R_m(T_3^4) = 2,\ R_m(T_4^4) = r_{m,4}(K)$ $R_m(T_5^4) = 2,\ R_m(T_6^4) = 2$

Table 3: Possible ranks for support orbits in $X_{2,k}/G_2$ with support sizes k for all fields K, with the exception $K = \mathbb{Z}/2\mathbb{Z}$ where $R_s(T_4^4) = R_m(T_4^4) = 3$

In Table 3, the functions dependent on the field K are defined as

$$r_{s,4}(K) := \begin{cases} 2 & \text{if } \operatorname{char} K \neq 2, \\ 3 & \text{if } \operatorname{char} K = 2, \end{cases}$$

$$r_{m,4}(K) := \begin{cases} 2 & \text{if } x^2 - A \in K[x] \text{ is reducible } \forall A \in K \text{ and } \operatorname{char} K \neq 2, \\ 3 & \text{otherwise.} \end{cases}$$

In particular, $r_{m,4}(\mathbb{R}) = 3$ and $r_{m,4}(\mathbb{C}) = 2$.

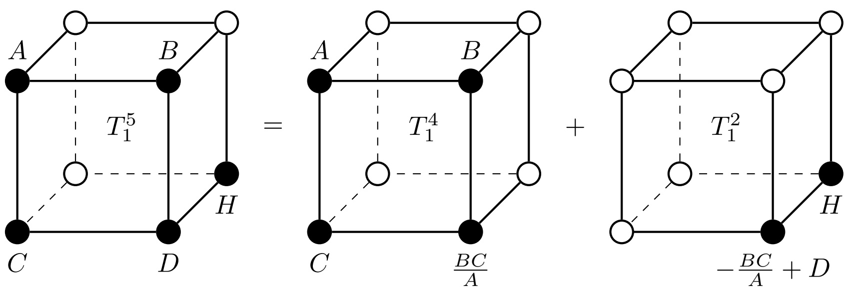
It is possible to construct an explicit field extension L from any starting field K with $\operatorname{char} K \neq 2$ such that $r_{m,4}(L) = 2$, without explicitly demanding that L should be the algebraic closure of K, even if sometimes L will actually be that closure, such as with the case $K = \mathbb{R}$. Starting with K and setting $K_0 := K$, construct for integers $i \geq 1$

$$K_i := K_{i-1}\left(\{\sqrt{A} : A \in K_{i-1}\}\right) = \operatorname{Span}\{\sqrt{A} : A \in K_{i-1}\}, \tag{34}$$

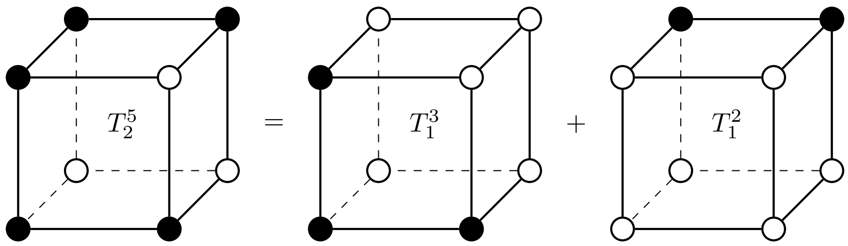
which in turn are fields if K_{i-1} is. This is because any two elements in K_i are linear combinations of $\{\sqrt{A_j}\}_{j=1}^m$ for some $\{A_j\}_{j=1}^m \in K_{i-1}$, and the two elements can be added, multiplied and have inverses in K_i as they can and have in $K_{i-1}\left(\{\sqrt{A_j}\}_{j=1}^m\right)$. The latter is a field due to being a finite chain of simple field extensions $K_{i-1} \subset K_{i-1}\left(\sqrt{A_1}\right) \subset \left(K_{i-1}\left(\sqrt{A_1}\right)\right)\left(\sqrt{A_2}\right)$ and so on. Then, let $L := \bigcup_{i=0}^{\infty} K_i$ which is again a field due to every two elements existing in a common K_i for large enough i since $K_i \subset K_j$ when $i \leq j$, where they can be added, multiplied and have inverses in $K_i \subset L$. This finally means that $r_{m,4}(L) = 2$ since any $A \in L$ must also lie in some K_i, where then the polynomial $x^2 - A$ have roots in $K_{i+1} \subset L$, making it reducible over L.

4.3.3 Support size $k = 5$

Moving on to support size five, the elements of T_1^5 will have rank two using a similar method as with T_3^4 in (25), shown as follows.



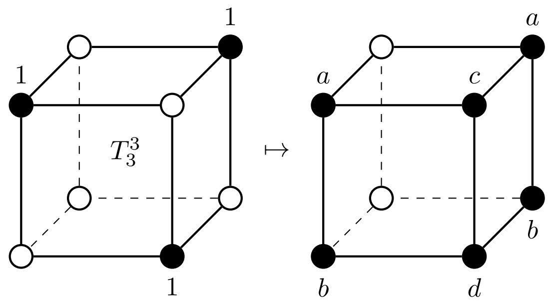
For the next orbit T_2^5 there are no rank larger than three due to all such tensors having the decomposition



Then, bijective linear maps are used on smaller support sizes, where if there is a t with tensor rank two or three that can be mapped to somewhere in T_2^5 then there exists a rank within T_2^5 equaling two or three. From the one-tensor in T_3^3, that is, the tensor with only ones or zeroes, the linear map

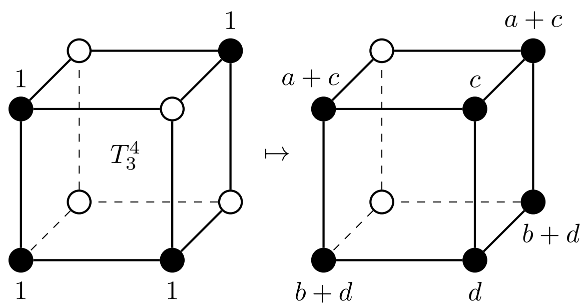
$$x_1 \mapsto ax_1 + bx_2,$$
$$x_2 \mapsto cx_1 + dx_2$$

sends it according to



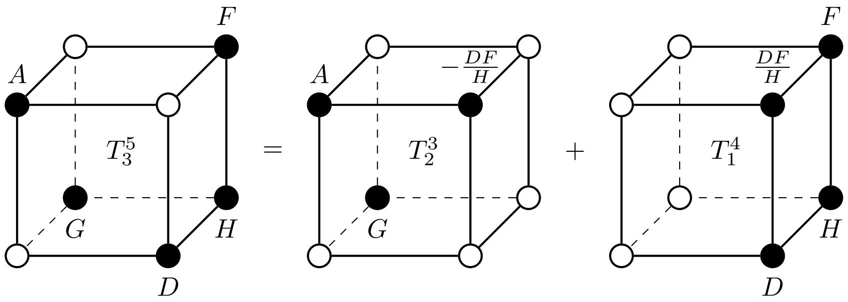
(35)

where $c = 0$ results in both it being mapped to T_2^5 and the map being bijective, and so there is a tensor in T_2^5 with rank three, due to $R(T_1^3) = 3$. The same type of map applied instead to the one-tensor of T_3^4 results in



where choosing $c = 0$ and $b + d \neq 0$ results in it being mapped to T_2^5 and the map being bijective, and so some tensor in T_2^5 has rank two, except for the field $K = \mathbb{Z}/2\mathbb{Z}$. But as seen before, the ranks of T_2^5 cannot be higher than three, which results in $R_s(T_2^5) = 2$, and $R_s(T_2^5) = 3$ when $K = \mathbb{Z}/2\mathbb{Z}$, and the maximum rank $R_m(T_2^5) = 3$.

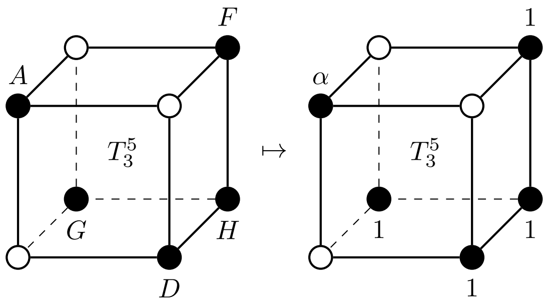
Similarly to T_4^4, the maximum rank of T_3^5 cannot be more than three, since tensors can always be written as



For each element T_3^5, there is again similarly to (27) a bijection

$$
\begin{aligned}
&Hx_2 \mapsto x_2,\ y_2 \mapsto y_2,\ z_2 \mapsto z_2,\\
&Fx_1 \mapsto y_2,\ x_1 \mapsto x_1,\\
&\frac{G}{H} y_1 \mapsto y_1,\\
&\frac{D}{H} z_1 \mapsto z_1
\end{aligned}
$$

such that



(36)

where $\alpha = \frac{AFH^2}{GD}$. Then, assuming such a tensor has rank two, it can be written as (28), with the only difference being the equation (31) having been changed to

$$(k_1 k_2 - k_1' k_2')\, b = 1. \tag{37}$$

Thus, using the same method as with T_4^4 results in $k_1' = k_2'$ and $k_1 = k_2$. Using (30) in (29), and (37) in (32), leads to

$$
\begin{aligned}
&\frac{a}{b} + (k_1')^2 \alpha = 1,\\
&\frac{a}{b} = -k_1' \alpha.
\end{aligned}
$$

This results in

$$(k_1')^2 - k_1' - \frac{1}{\alpha} = 0, \tag{38}$$

where, again, whether or not there are solutions depends on choices of field K and parameter α. For the other variable k_1, (37) and (38), and then (29), results in

$$k_1^2 = \frac{1}{b} + (k_1')^2 = \frac{1}{b} + k_1' + \frac{1}{\alpha} = k_1 + \frac{1}{\alpha}.$$

In other words, k_1 is another solution to (38), which has to be different to k_1' as in the case T_4^4. Given that two different k_1, k_1' exist that solve (38), b and a can be computed by inserting k_1, k_1' in (29) and (30).

Whether or not the rank of a tensor in T_3^5 will depend on on choices of field K and parameter α. However, in contrast to T_4^4, in every K with $\operatorname{char} K \neq 2$, there will be the case $\alpha = -4$ which results in a double root of (38), and so k_1 and k_1' have to be equal, which contradicts (37). Therefore, the maximum rank of T_3^5 will always be three for such fields K.

On the other hand, when $\operatorname{char} K = 2$, double roots will never occur. Existence of a double root β of a quadratic polynomial would mean that it is the solution to equation

$$0 = (x - \beta)^2 = x^2 - 2x + \beta^2 = x^2 + \beta^2,$$

which is not of the same form as (38) when $\operatorname{char} K = 2$. As such, when $\operatorname{char} K = 2$, the ranks will be two when (38) is solvable, and three when not. In $\operatorname{char} K = 2$, this is the same as

$$f_A(x) = x^2 + x + A \in K[x]$$

being either reducible or not for non-zero A, since α was arbitrary. For $K = \mathbb{Z}/2\mathbb{Z}$, $f_A(x)$ is irreducible due to the only case $A = 1$ not working. However, in any other such K, that is, a non-trivial extension of $\mathbb{Z}/2\mathbb{Z}$, there is an element β such that $\beta^2 + \beta \neq 0$. Otherwise, unique polynomial solutions in $K[x]$ would give $\beta = 0$ or $\beta = 1$. But then

$$(x + \beta)(x + \beta + 1) = x^2 + (2\beta + 1)x + \beta^2 + \beta = x^2 + x + \beta^2 + \beta = f_{\beta^2+\beta}(x)$$

and so there is a case where (38) is solvable without double roots. In other words, for $\operatorname{char} K = 2$, $R_s(T_3^5) = 2$ when $K \neq \mathbb{Z}/2\mathbb{Z}$ and $R_s(T_3^5) = 3$ when $K = \mathbb{Z}/2\mathbb{Z}$.

For the maximum rank of T_3^5 to be two in characteristic two, $f_A(x)$ would have to be reducible for every $A \in K$, but this can be stated more generally as

$$R_m(T_3^5) = \begin{cases} 3 & \text{if } \operatorname{char} K \neq 2, \\ 2 & \text{if } \operatorname{char} K = 2 \text{ and all quadratic polynomials of } K[x] \text{ are reducible}, \\ 3 & \text{otherwise.} \end{cases}$$

This is because every quadratic polynomial with linear coefficient $a \neq 0$ can be written using a change of variables $x = ay$ as

$$x^2 + ax + A = y^2a^2 + y^2a^2 + A = a^2\left(y^2 + y + A/a^2\right)$$

which is then reducible if and only if f_{A/a^2} is. As for quadratic polynomials without a linear coefficient, they are always reducible as long as char $K = 2$. This is because taking the square of an element

$$K \longrightarrow K,$$
$$r \longmapsto r^2$$

is a ring homomorphism in characteristic two, which is also injective due to only 0 being in its kernel. Thus, it is also surjective for finite K. If K is infinite, the field is also a field extension of $\mathbb{Z}/2\mathbb{Z}$, and then a vector space over $\mathbb{Z}/2\mathbb{Z}$. Then any element r of the infinite K is of the form $r = \sum_{i=1}^{N} \alpha_i$ and thus lies in the finite field $\mathbb{Z}/2\mathbb{Z}\left(\{\alpha_i\}_{i=1}^{N}\right) \subset K$. But then there is an $r' \in \mathbb{Z}/2\mathbb{Z}\left(\{\alpha_i\}_{i=1}^{N}\right) \subset K$ such that $(r')^2 = r$ since squaring is surjective over finite fields. This means that squaring is surjective and thus bijective even for infinite fields of characteristic two.

Going back to char $K \neq 2$, in most such K there is some choice of α such that there are two different k_1, k_1' that solve (38), since it amounts to finding an α such that

$$\{k_1, k_1'\} = \left\{\frac{1}{2} \pm \sqrt{\frac{1}{4} + \frac{1}{\alpha}}\right\}.$$

This is equivalent to $\frac{1}{4} + \frac{1}{\alpha}$ being a non-zero square in K. This is guaranteed to work in every such K with char $K \neq 2$ or 3, where for example 1 is the only non-zero square in $\mathbb{Z}/3\mathbb{Z}$, which would need $\frac{1}{4} + \frac{1}{\alpha} = 1$ which is the same as $1/\alpha = 0$ in characteristic three. For (38) to be solvable without double roots when char $K = 3$, the condition then becomes that there is some non-zero $A \in K \backslash \{-1\}$ such that $1 + A$ is square in K. But for every extension of $\mathbb{Z}/3\mathbb{Z}$, there must exist a $B \notin \mathbb{Z}/3\mathbb{Z}$ where then $B^2 \neq 1$, since it is known that the only two solutions to $x^2 - 1$ are ± 1. Then the choice $A = B^2 - 1$ gives a non-zero square $A + 1$, and so the only field with char $K \neq 3$ that results in support rank two is $K = \mathbb{Z}/3\mathbb{Z}$.

In summary, while ranks of individual tensors in T_3^5 are affected by choice of coefficients and field, the support rank and maximum rank of T_3^5 equal two and three respectively, independently of the field, assuming the characteristic is not two or three.

The updated table is now as follows, with the case $K = \mathbb{Z}/2\mathbb{Z}$ from here on being viewed separately as a rank denoted $R_2(\ldots)$, since it is the only field where the orbits as sets are singletons.

k	$\lvert X_{2,k}/G_2 \rvert$	Support ranks $R_s\left(T_i^k\right)$	Max ranks $R_m\left(T_i^k\right)$	Rank $R_2\left(T_i^k\right)$ in $\mathbb{Z}/2\mathbb{Z}$
0	1	1	1	1
1	1	1	1	1
2	3	$R_s\left(T_1^2\right)=1$ $R_s\left(T_2^2\right)=2$ $R_s\left(T_3^2\right)=2$	$R_m\left(T_1^2\right)=1$ $R_m\left(T_2^2\right)=2$ $R_m\left(T_3^2\right)=2$	$R_2\left(T_1^2\right)=1$ $R_2\left(T_2^2\right)=2$ $R_2\left(T_3^2\right)=2$
3	3	$R_s\left(T_1^3\right)=2$ $R_s\left(T_2^3\right)=2$ $R_s\left(T_3^3\right)=3$	$R_m\left(T_1^3\right)=2$ $R_m\left(T_2^3\right)=2$ $R_m\left(T_3^3\right)=3$	$R_2\left(T_1^3\right)=2$ $R_2\left(T_2^3\right)=2$ $R_2\left(T_3^3\right)=3$
4	6	$R_s\left(T_1^4\right)=1$ $R_s\left(T_2^4\right)=3$ $R_s\left(T_3^4\right)=2$ $R_s\left(T_4^4\right)=r_{s,4}\left(K\right)$ $R_s\left(T_5^4\right)=2$ $R_s\left(T_6^4\right)=2$	$R_m\left(T_1^4\right)=2$ $R_m\left(T_2^4\right)=3$ $R_m\left(T_3^4\right)=2$ $R_m\left(T_4^4\right)=r_{m,4}\left(K\right)$ $R_m\left(T_5^4\right)=2$ $R_m\left(T_6^4\right)=2$	$R_2\left(T_1^4\right)=1$ $R_2\left(T_2^4\right)=3$ $R_2\left(T_3^4\right)=2$ $R_2\left(T_4^4\right)=3$ $R_2\left(T_5^4\right)=2$ $R_2\left(T_6^4\right)=2$
5	3	$R_s\left(T_1^5\right)=2$ $R_s\left(T_2^5\right)=2$ $R_s\left(T_3^5\right)=r_{s,5}\left(K\right)$	$R_m\left(T_1^5\right)=2$ $R_m\left(T_2^5\right)=3$ $R_m\left(T_3^5\right)=r_{m,5}\left(K\right)$	$R_2\left(T_1^5\right)=2$ $R_2\left(T_2^5\right)=3$ $R_2\left(T_3^5\right)=3$

Table 4: Possible ranks for support orbits in $X_{2,k}/G_2$ with support sizes k for all fields K, with $K = \mathbb{Z}/2\mathbb{Z}$ being handled separately

In Table 4,

$$r_{s,5}\left(K\right) := \begin{cases} 3 & \text{if } K = \mathbb{Z}/2\mathbb{Z} \text{ or } K = \mathbb{Z}/3\mathbb{Z}, \\ 2 & \text{otherwise,} \end{cases}$$

$$r_{m,5}\left(K\right) := \begin{cases} 3 & \text{if } \operatorname{char} K \neq 2, \\ 2 & \text{if } \operatorname{char} K = 2 \text{ and all quadratic polynomials of } K[x] \text{ are reducible,} \\ 3 & \text{otherwise} \end{cases}$$

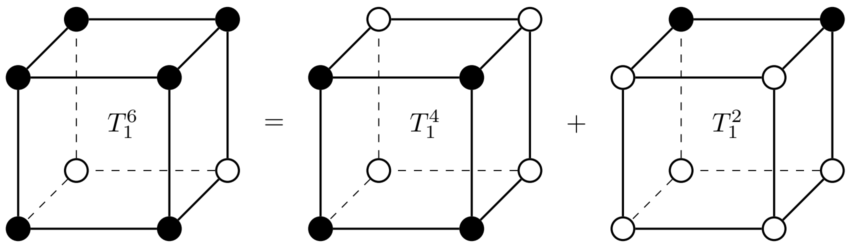
and in particular, $r_{m,5}\left(K\right) = 3$ when K is finite with $|K| = m$. This is because the number of quadratic monomials is strictly larger than the number of possible solution pairs. There are m^2 numbers of quadratic monomials, and there are $m + \binom{m}{2}$ possible solutions to them, with m double roots and $\binom{m}{2}$ separate roots. However, there is the strict inequality

$$m + \binom{m}{2} = m + \frac{m\left(m-1\right)}{2} < m + m\left(m-1\right) = m^2$$

for every $m \geq 2$. Then every monomial cannot have roots in K, and so the only candidates for maximum rank of two are infinite fields of characteristic two. It is possible to construct explicit field extensions L of any K with $\operatorname{char} K = 2$ such that $r_{m,5}\left(L\right) = 2$ by using a similar method as with $R_m\left(T_4^4\right)$ and (34), but using roots of quadratic monomials instead of just square roots $\sqrt{A}$. Similarly to that case, the extension L is not explicitly demanded to be the algebraic closure of K.

4.3.4 Support sizes $k = 6, 7, 8$

For the next support sizes six and seven, a similar method will be used as on T_2^5. In the case T_1^6, it is firstly clear that it cannot include ranks higher than three due to that a tensor in the orbit in general can be written in the decomposition



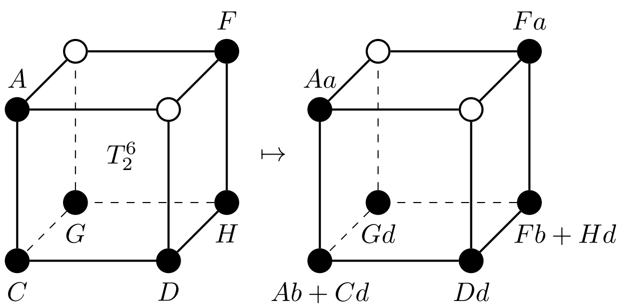
where in particular the one-tensor of T_1^6 has rank two. Rank three is also in T_1^6, seen by choosing map parameters a, b, c, d in (35) to be non-zero, resulting in the new tensor being in T_1^6, and such that the map is bijective, which can be done in every field but $\mathbb{Z}/2\mathbb{Z}$ by choosing $a = b = c = 1$ and $d \neq 1$.

This mapping method can be done in reverse as well, where the original tensor is of larger support size. The difference is that when mapping smaller supports to larger, it generally entails that there exists a specific rank in the orbit with the larger support, as seen above. In contrast, mapping larger supports to smaller will more generally give a list of the ranks the orbit with the larger support can possibly have.

For example, a general element of T_2^6 being mapped by

$$\begin{aligned} x_1 &\mapsto ax_1 + bx_2, \\ x_2 &\mapsto dx_2 \end{aligned} \tag{39}$$

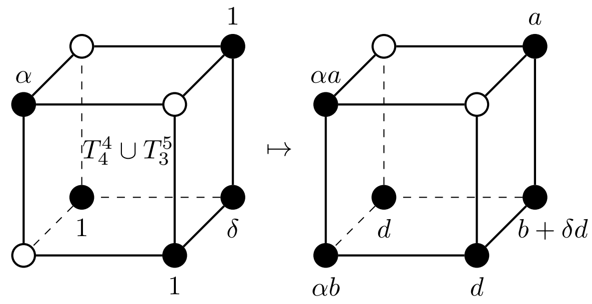
results in



(40)

Regardless of choices for non-zero A, C, D, F, G, H, there are choices for non-zero b, d such that one or both $Ab + Cd$ and $Fb + Hd$ are zero, and such the tensor being mapped to either T_4^4 or T_3^5. This means that the possible ranks are $R\left(T_2^6\right) \subset R\left(T_4^4\right) \cup R\left(T_3^5\right)$. On the other way around, a general element from T_4^4 or T_3^5, written using scaling as in (27) and (36) where here $\delta = 0$ if in T_4^4 and $\delta = 1$

if in T_3^5, is transformed by the map in (39) according to



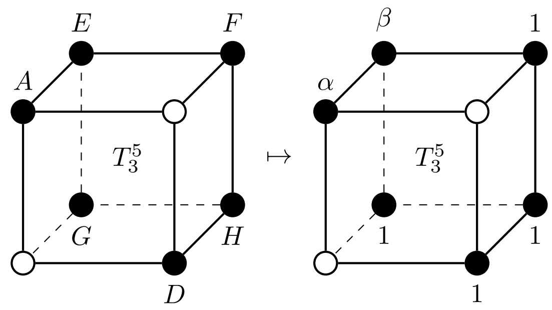
(41)

where b, d can be chosen such that $b + \delta d \neq 0$, that is, landing in T_2^6 regardless of the parameters δ, α, given that $K \neq \mathbb{Z}/2\mathbb{Z}$.

This finally results in $R\left(T_2^6\right) = R\left(T_4^4\right) \cup R\left(T_3^5\right)$ for all fields $K \neq \mathbb{Z}/2\mathbb{Z}$. In the case $K \neq \mathbb{Z}/2\mathbb{Z}$, the one-tensor, the then only element in T_2^6, is mapped according to (40) to T_4^4, and so has rank three as seen in Table 3, which also means $R\left(T_2^6\right) = R\left(T_4^4\right) \cup R\left(T_3^5\right) = \{3\}$ still holds. A priori, $R_m\left(T_2^6\right) = \max\{R_m\left(T_4^4\right), R_m\left(T_3^5\right)\}$ and $R_s\left(T_2^6\right)$ is the minimum, which is true, but can be simplified. From Table 4, $R_m\left(T_3^5\right) = 3$ only when the char $K = 2$, while then $R_m\left(T_4^4\right) = 3$. Thus $R_m\left(T_2^6\right) = 3$ for every field.

The next orbit T_3^6 is handled analogously to T_2^6, where the same mapping (39) takes all tensors of T_3^6 to either T_2^6 or T_3^5 by a similar argument, and so $R\left(T_3^6\right) \subset R\left(T_2^6\right) \cup R\left(T_3^5\right)$. The same mapping also takes tensors of T_1^7 to either T_2^6, T_3^6 or T_3^5, resulting in $R\left(T_1^7\right) \subset R\left(T_2^6\right) \cup R\left(T_3^5\right)$, where then in particular $R_m\left(T_1^7\right) \leq 3$.

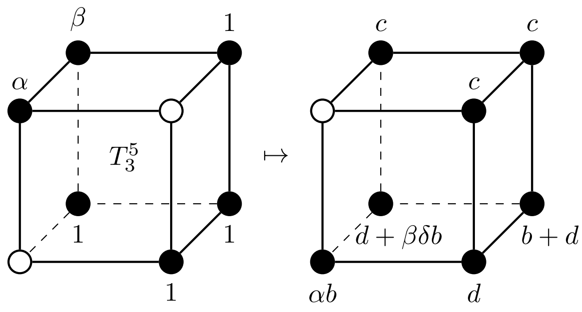
Moreover, tensors in T_3^6 can be scaled bijectively using the same mapping as in (36) on T_3^5. The difference is that the added non-zero vertex to T_3^6 compared with T_3^5 cannot be chosen to be 1, and will have two parameters α, β from the following scaling.



Similarly to (41), the mapping

$$x_1 \mapsto bx_2,$$
$$x_2 \mapsto cx_1 + dx_2$$

takes tensors from T_3^5 and T_2^6, where here $\delta = 0$ if in T_3^5 and $\delta = 1$ if in T_3^6, according to

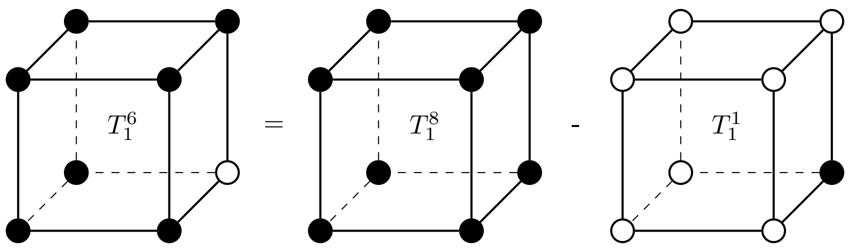


They are taken to T_3^6 by choosing b, d such that $b + d$ and $d + \beta\delta b$ are non-zero, which can be done regardless of α, β, δ when not in $\mathbb{Z}/2\mathbb{Z}$. Then,

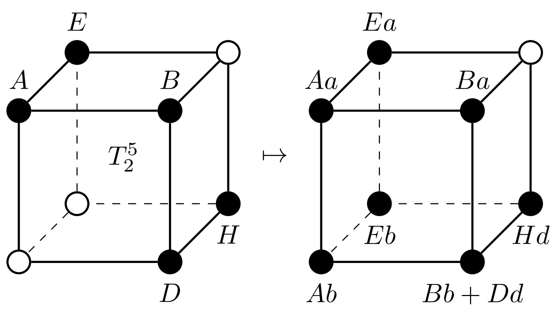
$$R\left(T_3^6\right) = R\left(T_2^6\right) \cup R\left(T_3^5\right) = R\left(T_2^6\right),$$

which holds for every field, since $R\left(T_2^6\right) \cup R\left(T_3^5\right) = \{3\}$ in $\mathbb{Z}/2\mathbb{Z}$.

Finishing up support size seven, $R_s\left(T_1^7\right) = 2$ for every field since it can be written as



and $R_m\left(T_1^7\right) = 3$ for every field but $\mathbb{Z}/2\mathbb{Z}$. This is because $R_m\left(T_1^7\right) \leq 3$, as shown earlier, and because tensors in T_2^5 can be mapped by (39) to $R_m\left(T_1^7\right) = 3$ according to



when choosing non-zero b, d such that $Bb + Dd \neq 0$. This is always possible when not in $\mathbb{Z}/2\mathbb{Z}$ and since $R_m\left(T_2^5\right) = 3$ when not in $\mathbb{Z}/2\mathbb{Z}$, $R_m\left(T_1^7\right) = 3$ in the same field.

For the final orbit T_1^8, the full support, its possible ranks are every rank that a smaller support can have. This is because, as stated earlier in Section 3.2.3, for each none-zero tensor there is an isomorphism turning it to a full support given that the field is not $\mathbb{Z}/2\mathbb{Z}$. On the other way around, for each tensor with full support there is an isomorphism turning at least one of the coefficients zero.

This means that $R\left(T_1^8\right) = \{1,2,3\}$ for every field $K \neq \mathbb{Z}/2\mathbb{Z}$, and $R\left(T_1^8\right) = 1$ when $K = \mathbb{Z}/2\mathbb{Z}$.

Every rank structure for every field K can now be summarized in Table 5

k	$\lvert X_{2,k}/G_2 \rvert$	Support ranks $R_s\left(T_i^k\right)$	Max ranks $R_m\left(T_i^k\right)$
0	1	1	1
1	1	1	1
2	3	$R_s\left(T_1^2\right) = 1,\ R_s\left(T_2^2\right) = 2$ $R_s\left(T_3^2\right) = 2$	$R_m\left(T_1^2\right) = 1,\ R_m\left(T_2^2\right) = 2$ $R_m\left(T_3^2\right) = 2$
3	3	$R_s\left(T_1^3\right) = 2,\ R_s\left(T_2^3\right) = 2$ $R_s\left(T_3^3\right) = 3$	$R_m\left(T_1^3\right) = 2,\ R_m\left(T_2^3\right) = 2$ $R_m\left(T_3^3\right) = 3$
4	6	$R_s\left(T_1^4\right) = 1,\ R_s\left(T_2^4\right) = 3$ $R_s\left(T_3^4\right) = 2,\ R_s\left(T_4^4\right) = r_{s,4}(K)$ $R_s\left(T_5^4\right) = 2,\ R_s\left(T_6^4\right) = 2$	$R_m\left(T_1^4\right) = 2,\ R_m\left(T_2^4\right) = 3$ $R_m\left(T_3^4\right) = 2,\ R_m\left(T_4^4\right) = r_{m,4}(K)$ $R_m\left(T_5^4\right) = 2,\ R_m\left(T_6^4\right) = 2$
5	3	$R_s\left(T_1^5\right) = 2,\ R_s\left(T_2^5\right) = 2$ $R_s\left(T_3^5\right) = r_{s,5}(K)$	$R_m\left(T_1^5\right) = 2,\ R_m\left(T_2^5\right) = 3$ $R_m\left(T_3^5\right) = r_{m,5}(K)$
6	3	$R_s\left(T_1^6\right) = 2$ $R_s\left(T_2^6\right) = \min\left\{R_s\left(T_4^4\right), R_s\left(T_3^5\right)\right\}$ $R_s\left(T_3^6\right) = R_s\left(T_2^6\right)$	$R_m\left(T_1^6\right) = 3$ $R_m\left(T_2^6\right) = 3$ $R_m\left(T_3^6\right) = 3$
7	1	2	3
8	1	1	3

Table 5: All rank structures for support classes in $K^2 \otimes K^2 \otimes K^2$ for all fields $K \neq \mathbb{Z}/2\mathbb{Z}$

with case functions

$$r_{s,4}(K) := \begin{cases} 2 & \text{if } \operatorname{char} K \neq 2, \\ 3 & \text{if } \operatorname{char} K = 2, \end{cases}$$

$$r_{m,4}(K) := \begin{cases} 2 & \text{if } x^2 - A \in K[x] \text{ is reducible } \forall A \in K \text{ and } \operatorname{char} K \neq 2, \\ 3 & \text{otherwise,} \end{cases}$$

$$r_{s,5}(K) := \begin{cases} 3 & \text{if } K = \mathbb{Z}/2\mathbb{Z} \text{ or } K = \mathbb{Z}/3\mathbb{Z}, \\ 2 & \text{otherwise,} \end{cases}$$

$$r_{m,5}(K) := \begin{cases} 3 & \text{if } \operatorname{char} K \neq 2, \\ 2 & \text{if } \operatorname{char} K = 2 \text{ and all quadratic polynomials of } K[x] \text{ are reducible,} \\ 3 & \text{otherwise.} \end{cases}$$

For particular fields $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}/2\mathbb{Z}$, the results can be seen in the following Tables 6, 7, 8 respectively.

k	$\lvert X_{2,k}/G_2 \rvert$	Support ranks in $\mathbb{R}$	Max ranks in $\mathbb{R}$
0	1	1	1
1	1	1	1
2	3	$R_s\left(T_1^2\right)=1,\ R_s\left(T_2^2\right)=2$ $R_s\left(T_3^2\right)=2$	$R_m\left(T_1^2\right)=1,\ R_m\left(T_2^2\right)=2$ $R_m\left(T_3^2\right)=2$
3	3	$R_s\left(T_1^3\right)=2,\ R_s\left(T_2^3\right)=2$ $R_s\left(T_3^3\right)=3$	$R_m\left(T_1^3\right)=2,\ R_m\left(T_2^3\right)=2$ $R_m\left(T_3^3\right)=3$
4	6	$R_s\left(T_1^4\right)=1,\ R_s\left(T_2^4\right)=3$ $R_s\left(T_3^4\right)=2,\ R_s\left(T_4^4\right)=2$ $R_s\left(T_5^4\right)=2,\ R_s\left(T_6^4\right)=2$	$R_m\left(T_1^4\right)=2,\ R_m\left(T_2^4\right)=3$ $R_m\left(T_3^4\right)=2,\ R_m\left(T_4^4\right)=3$ $R_m\left(T_5^4\right)=2,\ R_m\left(T_6^4\right)=2$
5	3	$R_s\left(T_1^5\right)=2,\ R_s\left(T_2^5\right)=2$ $R_s\left(T_3^5\right)=2$	$R_m\left(T_1^5\right)=2,\ R_m\left(T_2^5\right)=3$ $R_m\left(T_3^5\right)=3$
6	3	$R_s\left(T_1^6\right)=2,\ R_s\left(T_2^6\right)=2$ $R_s\left(T_3^6\right)=2$	$R_m\left(T_1^6\right)=3\ R_m\left(T_2^6\right)=3$ $R_m\left(T_3^5\right)=3$
7	1	2	3
8	1	1	3

Table 6: All rank structures for support classes in $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$

k	$\lvert X_{2,k}/G_2 \rvert$	Support ranks in $\mathbb{C}$	Max ranks in $\mathbb{C}$
0	1	1	1
1	1	1	1
2	3	$R_s\left(T_1^2\right)=1,\ R_s\left(T_2^2\right)=2$ $R_s\left(T_3^2\right)=2$	$R_m\left(T_1^2\right)=1,\ R_m\left(T_2^2\right)=2$ $R_m\left(T_3^2\right)=2$
3	3	$R_s\left(T_1^3\right)=2,\ R_s\left(T_2^3\right)=2$ $R_s\left(T_3^3\right)=3$	$R_m\left(T_1^3\right)=2,\ R_m\left(T_2^3\right)=2$ $R_m\left(T_3^3\right)=3$
4	6	$R_s\left(T_1^4\right)=1,\ R_s\left(T_2^4\right)=3$ $R_s\left(T_3^4\right)=2,\ R_s\left(T_4^4\right)=2$ $R_s\left(T_5^4\right)=2,\ R_s\left(T_6^4\right)=2$	$R_m\left(T_1^4\right)=2,\ R_m\left(T_2^4\right)=3$ $R_m\left(T_3^4\right)=2,\ R_m\left(T_4^4\right)=2$ $R_m\left(T_5^4\right)=2,\ R_m\left(T_6^4\right)=2$
5	3	$R_s\left(T_1^5\right)=2,\ R_s\left(T_2^5\right)=2$ $R_s\left(T_3^5\right)=2$	$R_m\left(T_1^5\right)=2,\ R_m\left(T_2^5\right)=3$ $R_m\left(T_3^5\right)=3$
6	3	$R_s\left(T_1^6\right)=2,\ R_s\left(T_2^6\right)=2$ $R_s\left(T_3^6\right)=2$	$R_m\left(T_1^6\right)=3\ R_m\left(T_2^6\right)=3$ $R_m\left(T_3^6\right)=3$
7	1	2	3
8	1	1	3

Table 7: All rank structures for support classes in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$

k	$\|X_{2,k}/G_2\|$	Ranks $R_2\left(T_i^k\right)$ in $\mathbb{Z}/2\mathbb{Z}$
0	1	1
1	1	1
2	3	$R_2\left(T_1^2\right)=1,\ R_2\left(T_2^2\right)=2$ $R_2\left(T_3^2\right)=2$
3	3	$R_2\left(T_1^3\right)=2,\ R_2\left(T_2^3\right)=2$ $R_2\left(T_3^3\right)=3$
4	6	$R_2\left(T_1^4\right)=1,\ R_2\left(T_2^4\right)=3$ $R_2\left(T_3^4\right)=2,\ R_2\left(T_4^4\right)=3$ $R_2\left(T_5^4\right)=2,\ R_2\left(T_6^4\right)=2$
5	3	$R_2\left(T_1^5\right)=2,\ R_2\left(T_2^5\right)=3$ $R_2\left(T_3^5\right)=3$
6	3	$R_2\left(T_1^6\right)=2,\ R_2\left(T_2^6\right)=3$ $R_2\left(T_3^6\right)=3$
7	1	2
8	1	1

Table 8: All rank structures for support classes in $K^2 \otimes K^2 \otimes K^2$ for $K = \mathbb{Z}/2\mathbb{Z}$

4.4 Support orbits for $n > 2$

For higher $n > 2$, it should be noted that smaller support sizes, specifically smaller or equal to $n-1$, will not give any new information about rank structures that is not included in the rank structures in dimension $n-1$. This is because whenever a support class has all its non-zero elements in the subspace

$$V_{n-1} := \text{Span}\{x_i \otimes y_j \otimes z_k\}_{i,j,k=1}^{n-1} \subset K^n \otimes K^n \otimes K^n,$$

it will have the same rank structure as the corresponding support class in the isomorphic $K^{n-1} \otimes K^{n-1} \otimes K^{n-1}$. If they are not all in V_{n-1}, then a process analogue to permuting rows and columns in a matrix can be used.

If the support size is smaller or equal to $n-1$ and a support class has a non-zero element a outside of V_{n-1}, for example being in the x_n-slice of the tensor, then there is some x_i-slice in V_{n-1} which has only zeroes. Then a permutation automorphism $x_i \to x_n$ will not move any elements already in V_{n-1} to outside of it. If after the permutation a is not in V_{n-1}, then repeat the process but with y- or z-variables instead, until $a \in V_{n-1}$. This can be repeated until all non-zero elements are in V_{n-1}, and so the rank structure of the support class is the same as its rank structure within the smaller subspace V_{n-1}.

From this, it can be seen that the first support size k for which the support rank of an orbit is not its only rank will be $k = 4$, regardless of n. This follows from the fact tensors in an orbit of support size three are equivalent, seen using scaling of x, y, z-variables. Starting from the one-tensor

$$t(1) = x_i y_j z_k + x_{i'} y_{j'} z_{k'} + x_{i''} y_{j''} z_{k''},$$

of such an orbit with three distinct terms, scale a variable $x_i \to Ax_i$ if $i \neq i'$ leaving y_j and z_k unchanged. Then scale $x_{i'} \to Bx_{i'}$. Finally, if $i' \neq i''$, scale $x_{i''} \to Cx_{i''}$, and otherwise scale the variable $y_{j''}$ or $z_{k''}$ with C since either $j' \neq j''$ or $k' \neq k''$ must hold. If from the start $i = i'$, then do similar scaling on the y or z-variables since either $j \neq j'$ or $k \neq k'$ must hold. This results in $t(1)$ being mapped by a bijective linear transformation to any tensor $Ax_iy_jz_k + Bx_{i'}y_{j'}z_{k'} + Cx_{i''}y_{j''}z_{k''}$ with non-zero A, B, C, and so tensors in the support orbit are all equivalent.

5 Summary and discussion

5.1 Conclusions regarding rank structures in $K^2 \otimes K^2 \otimes K^2$

Firstly, one takeaway is that support rank of an orbit is affected by the field only if that field has characteristic two and three. This occurs on six different support orbits. It should have been clear even before classifying the rank structure that the field $K = \mathbb{Z}/2\mathbb{Z}$ is an important case for support rank, as it is the only field for which support orbits are singletons, which of course means a change in rank, compared to another field, will always result in a change in support rank.

The maximum rank of the support orbits are instead dependent, with the exception of $\mathbb{Z}/2\mathbb{Z}$, on the degree to how much the underlying field is algebraic closed, where fields with a higher density of reducible second degree polynomials will have more tensors with lower rank. The difference between R_s and R_m becomes larger for more support orbits the larger the support size gets, as hypothesized in the beginning of Section 4. One explanation is that larger support sizes leads to more parameters for the tensors in an orbit, increasing the difficulty of finding linear maps transforming all tensors in an orbit to one another, in other words be equivalent. It should however be mentioned that two tensors having the same rank does not mean that they are equivalent.

Still, the only orbit where the number possible ranks are three is the full support, so it may be inconclusive to what degree the support size affects the span of ranks of an orbit, where perhaps the number of ranks would increase dramatically just as the support becomes full, instead of steadily increasing as the support size does. Better results in this regard could be obtained from higher n, as the pool of ranks in $K^2 \otimes K^2 \otimes K^2$ is only $\{0, 1, 2, 3\}$, and the maximum rank of the entire tensor product increases as n does, as seen in Proposition 2.

The symmetry between orbits of support sizes k and $8-k$ does not in general also appear in the rank structures, as questioned in the beginning of Section 4. Specifically, $R(T_i^k) = R(T_i^{8-k})$ does not hold for all k and i. One interesting results is that there is only one example of a field K which includes some symmetry of ranks, albeit not $R(T_i^k) = R(T_i^{8-k})$. For $K = \mathbb{Z}/3\mathbb{Z}$, equalities within the set $\{R_s(T_i^k)\}_i$ will follow for the corresponding orbits within $\{R_s(T_i^{8-k})\}_i$. In other words,

$$R_s(T_i^k) = R_s(T_j^k) \text{ if and only if } R_s(T_i^{8-k}) = R_s(T_j^{8-k}),$$

while similar does not hold for maximum or possible ranks even when $K = \mathbb{Z}/3\mathbb{Z}$.

5.2 Summary and further developments for support rank structures

Classifying rank structures among support classes can be a hard endeavour given that the number of support classes of size k is $\binom{n^3}{k}$, which will generally be very large even for small integers n. While Burnside's lemma can dramatically decrease the number of different rank structures by sieving through the support classes, there is the catch that the size of the groups of permutation automorphisms $|G_n|$ increases by $3!(n!)^3$, which is much faster than $\binom{n^3}{k}$ that had a growth which was already troubling. Even though the permutation automorphisms, the support classes, and the ways the former act on the latter, are finite, this would mean that attempts at brute forcing Burnside's lemma would be highly inefficient even for computers, albeit still theoretically possible.

Instead further work could be made of classifying the group G_n and its conjugacy classes for higher n. The group G_n changes, however, in a qualitative and geometric sense when $n > 2$. While support classes in $K^n \otimes K^n \otimes K^n$ can still be interpreted as geometric objects made out of n^3 small cubes, G_n will not preserve distances within the object for $n > 2$, in contrast to G_2. Thus, the group elements will in general not be geometric automorphisms on the cubes.

5.3 Final discussions on matrix multiplication rank

While there have been great developments in bounding tensor ranks and specifically bounding the theoretically best time complexity for matrix multiplication algorithms, in comparison to ordinary matrix rank there is still no general method of computing or tightly bounding tensor ranks. The new relaxations of rank introduced has contributed greatly to progress of bounding ordinary rank, and have often been introduced specifically to study the matrix multiplication tensor. One idea can be to not only continue to focus on border rank and support rank for general tensors, but to introduce other rank relaxations to specifically study other important bilinear algorithms. Those could then be generalized like border rank and support rank has been. This could coincide with finding new interesting families of tensors, since as mentioned before the family of Coppersmith-Winograd tensors has dominated for the last thirty years in terms of bounding the exponent of matrix multiplication ω. With the arguably most intuitive and yet very broad methods of using degenerations and linear maps to transform tensors in order to bound ranks, it is not possible to reach $\omega \leq 2$ using only Coppersmith-Winograd tensors. Thus, the best option, even though it may not be even close to easy, is to find other low-rank families of tensors to compare with matrix multiplication or other bilinear algorithms.

References

[1] Joshua A. Grochow and Cristopher Moore. "Designing Strassen's algorithm". In: *CoRR* (2017). arXiv: 1708.09398. URL: http://arxiv.org/abs/1708.09398.

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

[3] Thomas W. Judson. *Abstract Algebra: Theory and Applications*. Orthogonal Publishing, July 2019. ISBN: 978-1-944325-09-1.

[4] 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.

[5] Thomas D. Howell. "Global properties of tensor rank". In: *Linear Algebra and its Applications* 22 (1978), pp. 9–23. ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(78)90052-6. URL: https://www.sciencedirect.com/science/article/pii/0024379578900526.

[6] J. M. Landsberg. *Tensors : Geometry and Applications*. eng. Graduate studies in mathematics ; v. 128. Providence, R.I: American Mathematical Society, 2011. ISBN: 978-0-8218-6907-9.

[7] Markus Bläser, Matthias Christandl, and Jeroen Zuiddam. "The border support rank of two-by-two matrix multiplication is seven". In: *CoRR* (2017). arXiv: 1705.09652. URL: http://arxiv.org/abs/1705.09652.

[8] Markus Bläser. *Fast Matrix Multiplication*. Graduate Surveys 5. Theory of Computing Library, 2013, pp. 1–60. DOI: 10.4086/toc.gs.2013.005.

[9] Josh Alman. "Limits on the Universal Method for Matrix Multiplication". In: *Theory of Computing* 17.1 (2021), pp. 1–30. DOI: 10.4086/toc.2021.v017a001. URL: https://theoryofcomputing.org/articles/v017a001.

[10] Peter Bürgisser, Michael Clausen, and Amin Shokrollahi. *Algebraic Complexity Theory*. Springer, Jan. 1997. ISBN: 978-3-642-08228-3. DOI: 10.1007/978-3-662-03338-8.

[11] Henry Cohn and Christopher Umans. "Fast matrix multiplication using coherent configurations". In: *Proc. 24th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'13)* (2013), pp. 1074–1086. DOI: 10.1137/1.9781611973105.77.

[12] Alhussein Fawzi et al. "Discovering faster matrix multiplication algorithms with reinforcement learning". In: *Nature* 610 (Oct. 2022), pp. 47–53. DOI: 10.1038/s41586-022-05172-4.

[13] I. Martin Isaacs. *Character Theory of Finite Groups*. Dover, 1994. ISBN: 978-0-486-68014-9.

TRITA – SCI-GRU 2023:453