

Algebraic Methods for Tensor Data

Neriman Tokcan^{1,2}, Jonathan Gryak³, Kayvan Najarian^{3,4,5}, and Harm Derksen^{1,5}

Abstract. We develop algebraic methods for computations with tensor data. We give 3 applications: extracting features that are invariant under the orthogonal symmetries in each of the modes, approximation of the tensor spectral norm, and amplification of low rank tensor structure. We introduce colored Brauer diagrams, which are used for algebraic computations and in analyzing their computational complexity. We present numerical experiments whose results show that the performance of the alternating least square algorithm for the low rank approximation of tensors can be improved using tensor amplification.

Key words. tensors, Brauer diagrams, representation theory, invariant theory.

AMS subject classifications. 15A72, 15A69, 62-07, 22E45, 20G05

1. Introduction. Data in applications often is structured in higher dimensional arrays. Arrays of dimension d are also called d -way tensors, or *tensors of order d* . It is challenging to generalize methods for matrices, which are 2-dimensional arrays, to tensors of order 3 or higher. The notion of rank can be generalized from matrices to higher order tensors (see [13, 14]). Also, the spectral and nuclear norms are not only defined for matrices, but also for tensors of order ≥ 3 ([9, 27]). However, the rank, spectral norm, and nuclear norm of a higher order tensor are difficult to compute. In fact, the related decision problems are NP-complete. This was proved for the tensor rank in [10, 11], for the spectral norm in [12] and for the nuclear norm in [7].

We will use algebraic methods from classical invariant theory to perform various computations with tensors and analyze the computational complexity. Our methods are based on the description of tensor invariants of the orthogonal group by *Brauer diagrams* ([2, 8, 31]). Brauer diagrams are perfect matching graphs. We will discuss the background on Classical Invariant Theory and Brauer diagrams in Section 2. We will restrict ourselves to 3-way tensors. The techniques generalize to tensors of order ≥ 4 , but some of the formulas become more complicated. To perform computations with 3-way tensors, we generalize the notion of Brauer diagrams to colored trivalent graphs called *colored Brauer diagrams*, Section 3.

In this paper we consider 3 applications of our algebraic approach, namely invariant tensor features from data, approximations of the spectral and nuclear norm, and tensor amplification. In Subsection 4.1, we introduce the norm $\|\mathcal{T}\|_{\sigma,m}^m$ for $m \in \mathbb{N}$ to approximate the spectral norm of 3-way tensors. In Subsection 4.2, we show that $\|\mathcal{T}\|_{\sigma,2}$ is equal to the Euclidean norm (alternatively called Frobenius, Hilbert-Schmidt norm). In Subsection 4.4, we introduce an-

¹Department of Mathematics, University of Michigan, Ann Arbor

²The Eli and Edythe L. Broad Institute of MIT and Harvard, Cambridge, Massachusetts

³Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor

⁴Department of Emergency Medicine, University of Michigan, Ann Arbor

⁵Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor

other norm $\|T\|_{\#}$ that approximates the spectral norm. The main results are explicit formulas of these norms in terms of colored Brauer diagrams (see Theorem 4.2 and Proposition 4.5) and a comparison between the spectral norm and these approximations (see Proposition 4.8) is given. In Section 5, we study the low rank amplification methods based on the approximations of the spectral norm. We employ these amplification methods to obtain better initial guesses for the CP-ALS method; an algorithm for the low rank approximation to 3-way tensors is given (Section 5.3, Algorithm 5.1). In Section 6, we compare the ALS tensor approximation based on tensor amplification initialization with random initialization. In our experiments, we see that methods introduced in Section 5.3 give low rank r approximations ($r=1$ in Subsection 6.1 and $r=2$ in Subsection 6.2) with better fits and improved time efficiency compared to CP-ALS method.

1.1. Notation and Preliminaries. We will introduce the basic concepts and notation, which will lay the foundation for the rest of the paper. We will borrow most of our notation from [7] and [15].

As we have stated before, tensors are multi-dimensional arrays. The *order* of a tensor is the number of its dimensions (ways, modes). Vectors are tensors of order 1 and matrices are tensors of order 2. We will refer to tensors of order 3 or higher as higher-order tensors. Vectors are denoted by lower case letters $x \in \mathbb{R}^p$, matrices are denoted by capital letters $X \in \mathbb{R}^{p \times q}$, and higher-order tensors are denoted by capital calligraphic letters $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$. The (i_1, i_2, \dots, i_d) -th entry of the d -th order (d -way) tensor \mathcal{X} is denoted by $x_{i_1 i_2 \dots i_d}$.

The *vector outer product* of $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$ is denoted by $u \otimes v$ and it can be given as the matrix multiplication $uv^T \in \mathbb{R}^{p \times q}$.

The *inner product* of two same size tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$ is defined as follows:

$$(1.1) \quad \mathcal{X} \cdot \mathcal{Y} = \sum_{i_1=1}^{p_1} \sum_{i_2=1}^{p_2} \dots \sum_{i_d=1}^{p_d} x_{i_1 i_2 \dots i_d} y_{i_1 i_2 \dots i_d} \in \mathbb{R}.$$

It follows immediately that the norm of a tensor is the square root of the sum of the squares of all its elements:

$$(1.2) \quad \|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{p_1} \sum_{i_2=1}^{p_2} \dots \sum_{i_d=1}^{p_d} x_{i_1 i_2 \dots i_d}^2}.$$

This is analogous to the matrix Frobenius norm, see Section 1.3 for more details on the Frobenius norm.

A tensor $\mathcal{S} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$ is *rank one* if it can be written as an outer product of n vectors, i.e., $\mathcal{S} = u_1 \otimes u_2 \otimes \dots \otimes u_d$, $u_i \neq 0 \in \mathbb{R}^{p_i}$, $1 \leq i \leq d$. Such rank one tensors are also called *simple* or *pure*.

The best rank 1 approximation problem of a tensor $\mathcal{T} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$ can be stated as follows:

$$(1.3) \quad \min_{\mathcal{S}} \|\mathcal{T} - \mathcal{S}\| \text{ where } \mathcal{S} \text{ is a rank one tensor in } \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}.$$

The best rank 1 approximation problem is well-posed and NP-hard ([17, 28]). Different algebraic tools and algorithms have been proposed to find the global minimum of Problem 1.3 (see [28, 34]).

A tensor $\mathcal{S} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$ can be represented as a linear combination of rank 1 tensors:

$$(1.4) \quad \mathcal{S} = \sum_{i=1}^r \lambda_i u_{1,i} \otimes u_{2,i} \otimes \dots \otimes u_{d,i}$$

where r is sufficiently large, $\lambda_i \in \mathbb{R}$ and $u_{j,i} \in \mathbb{R}^{p_j}$ for $1 \leq i \leq r$, $1 \leq j \leq d$. The smallest such integer r is called the *rank* (*real rank*) of the tensor. The decomposition given in (1.4) is often referred to as the *rank r decomposition*, *CP* (*Candecom* / *Parafac*), or *Canonical Polyadic decomposition*. Let $U^{(j)} = [u_{j,1} u_{j,2} \dots u_{j,r}] \in \mathbb{R}^{p_j \times r}$, $1 \leq j \leq d$. We call these matrices as *factor matrices*. Then CP decomposition factorizes a d -way tensor into d factor matrices and a vector $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_r] \in \mathbb{R}^r$. The decomposition in (1.4) can be concisely expressed as $\mathcal{S} = \llbracket \Lambda ; U^{(1)}, U^{(2)}, \dots, U^{(d)} \rrbracket$. As in (1.3), the best rank r approximation problem for a tensor $\mathcal{T} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$ can be given as:

$$(1.5) \quad \min_{\Lambda, U^{(1)}, \dots, U^{(d)}} \|\mathcal{T} - \mathcal{S}\| \text{ where } \mathcal{S} = \llbracket \Lambda ; U^{(1)}, U^{(2)}, \dots, U^{(d)} \rrbracket.$$

The solution to Problem (1.5) does not always exist ([15, 29]). Alternating Least Squares (ALS) is the most common method used for the low rank approximation, since it is simple and easy to implement. However, it has some limitations: convergence is slow for some cases, it is heavily dependent on the initial guess of the factor matrices, and it may not converge to a global minimum (see [15, 17] for more details on the CP decomposition and ALS method). More details on the ALS algorithm for the low rank approximation are given in Section 5.3.

1.2. Invariant tensor features from data. Let $O_p(\mathbb{R})$ be the group of orthogonal $p \times p$ matrices. The group $O_p(\mathbb{R}) \times O_q(\mathbb{R})$ acts on the space $\mathbb{R}^{p \times q}$ of $p \times q$ matrices by left and right multiplication. A group element $(B, C) \in O_p(\mathbb{R}) \times O_q(\mathbb{R})$ acts on a matrix $A \in \mathbb{R}^{p \times q}$ by $(B, C) \cdot A = BAC^{-1}$. The singular values $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_r(A) \geq 0$ of a $p \times q$ matrix A are the features that are invariant under the actions of $O_p(\mathbb{R})$ and $O_q(\mathbb{R})$ on the rows and columns respectively. In other words, if B and C are orthogonal matrices, then $\lambda_i(BAC^{-1}) = \lambda_i(A)$ for $1 \leq i \leq r$. The function t_k given by $t_k(A) = \text{trace}((AA^T)^k) = \lambda_1(A)^{2k} + \lambda_2(A)^{2k} + \dots + \lambda_r(A)^{2k}$, $k \geq 0$ is also invariant under the actions of $O_p(\mathbb{R})$ and $O_q(\mathbb{R})$. The invariant functions $t_1(A), t_2(A), \dots$ are polynomials in the entries of the matrix A . It is known that the set $\{t_k : k \geq 0\}$ generates the ring of polynomial invariants under the action of $O_p(\mathbb{R}) \times O_q(\mathbb{R})$ on $p \times q$ matrices (see for example [8, §12.4.3, type **BDI**], but here we do not need any Pfaffians because we consider orthogonal groups and not special orthogonal groups). We will consider invariant features for 3-way tensors. Using classical invariant theory for the orthogonal group, we will describe polynomial tensor invariants in terms of colored Brauer diagrams. A similar approach to describing tensor invariants of orthogonal group actions can be found in the thesis [32, §4.2]. We will introduce the colored Brauer diagrams in Section 3 and use them to construct polynomial tensor invariants.

1.3. Approximations of the spectral and nuclear norm. Important norms on the space of $p \times q$ matrices are the Frobenius norm (or the Euclidean ℓ_2 -norm), the spectral norm (or operator norm), and the nuclear norm. These norms can be expressed in terms of the singular values of a matrix. If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$ are the singular values of a matrix A , then the Frobenius norm is $\|A\| = \|A\|_F = \sqrt{\lambda_1^2 + \lambda_2^2 + \dots + \lambda_r^2}$, the spectral norm is $\|A\|_\sigma = \lambda_1$, and the nuclear norm is $\|A\|_* = \lambda_1 + \lambda_2 + \dots + \lambda_r$. The nuclear norm can be seen as a convex relaxation of the rank of a matrix. It is used for example in some algorithms for the matrix completion problem which asks to complete a partially filled matrix such that the rank of the resulting matrix has minimal rank ([4, 5]). This problem has applications to collaborative filtering (see [26]). The spectral and nuclear norms generalize to higher order tensors.

Let \mathcal{T} be a tensor of order d in $\mathbb{R}^{p_1 \times p_2 \times \dots \times p_d}$. We define its *spectral norm* by

$$(1.6) \quad \|\mathcal{T}\|_\sigma = \sup \{ |\mathcal{T} \cdot u_1 \otimes u_2 \otimes \dots \otimes u_d| : u_j \in \mathbb{R}^{p_j}, \|u_j\| = 1, 1 \leq j \leq d \}.$$

It is known that the dual of the spectral norm is the *nuclear norm* and it can be defined as

$$(1.7) \quad \|\mathcal{T}\|_* = \inf \{ \sum_{i=1}^r |\lambda_i| : \mathcal{T} = \sum_{i=1}^r \lambda_i u_{1,i} \otimes u_{2,i} \otimes \dots \otimes u_{d,i}, \text{ where } \lambda_i \in \mathbb{R}, \\ u_{j,i} \in \mathbb{R}^{p_j}, \|u_{j,i}\| = 1, 1 \leq j \leq d, 1 \leq i \leq r \}.$$

These generalizations are more difficult to compute, as the corresponding decision problems are NP-hard ([7, 12]). As in the matrix case, nuclear norm of a tensor is considered as a convex relaxation of the tensor rank [6]. The nuclear and spectral norms of tensors play an important role in tensor completion problems [33]. Different methods to estimate and to evaluate the spectral norm and the nuclear norm and their upper and lower bounds have been studied by several authors (see [16, 20, 21, 25]).

The spectral norm is related to rank 1 approximation of a given tensor. If \mathcal{S} is a best rank 1 approximation of a given tensor \mathcal{T} , then $\|\mathcal{T} - \mathcal{S}\| = \sqrt{\|\mathcal{T}\|^2 - \|\mathcal{T}\|_\sigma^2}$, (Proposition 1.1, [21]).

We will give approximations of the spectral norm that can be computed in polynomial time using colored Brauer diagrams in Section 4. For every even d we define a norm $\|\cdot\|_{\sigma,d}$ that approximates the spectral norm $\|\cdot\|_\sigma$ such that $\|\cdot\|_{\sigma,d}^d$ is a polynomial function of degree d and $\lim_{d \rightarrow \infty} \|\mathcal{T}\|_{\sigma,d} = \|\mathcal{T}\|_\sigma$ for any tensor \mathcal{T} . One of our main results is an explicit formula for the norm $\|\cdot\|_{\sigma,4}$ for tensors of order 3 in terms of colored Brauer diagrams (see Theorem 4.2), which allows us to compute this norm efficiently. We also introduce another norm $\|\cdot\|_\#$ (see Definition 4.3, Proposition 4.5) and show that it is, in some sense, a better approximation to the spectral norm than $\|\cdot\|_{\sigma,4}$ (see Proposition 4.8).

$\|\cdot\|$ will stand for the Frobenius norm throughout the paper.

1.4. Tensor amplification. If A is a real matrix with singular values $\lambda_1, \dots, \lambda_r$, then the matrix $AA^T A$ has singular values $\lambda_1^3, \lambda_2^3, \dots, \lambda_r^3$. The map $A \mapsto AA^T A$ has the effect of amplifying the low rank structure corresponding to larger singular values, while suppressing the smaller singular values that typically correspond to noise. Using colored Brauer diagrams, we will construct similar amplification maps for tensors of order 3 in Section 5. We also will present numerical experiments whose results show that tensor amplification can reduce the running time of the alternating least square algorithm for the low rank tensor approximation, while producing a better approximation.

2. Brauer diagrams. In this section, we will give an overview of the classical invariant theory of the orthogonal group. We recall the relation between Brauer diagrams and invariant tensors for the orthogonal group.

2.1. Orthogonal transformations on tensors. Let $V \cong \mathbb{R}^n$ be a Euclidean vector space with basis e_1, e_2, \dots, e_n . The orthogonal group $O(V) = O_n(\mathbb{R}) = \{A \in \mathbb{R}^{n \times n} \mid AA^T = I\}$ acts on V . On V we have an inner product that allows us to identify V with its dual space V^* . We consider the d -fold tensor product of V :

$$(2.1) \quad V^{\otimes d} = \underbrace{V \otimes V \otimes \dots \otimes V}_d \cong \mathbb{R}^{n \times n \times \dots \times n} \cong \mathbb{R}^{n^d}.$$

There are various ways to think of elements of $V^{\otimes d}$. The following statement is well known (Chapter 2, [17]).

Lemma 2.1. *There are bijections between the following sets:*

1. the set of **tensors** $V^{\otimes d}$;
2. $(V^{\otimes d})^*$, the set of **linear maps** $V^{\otimes d} \rightarrow \mathbb{R}$;
3. the set of **\mathbb{R} -multilinear maps** $V^d \rightarrow \mathbb{R}$.

Proof. We have a multi-linear map $\iota : V^d \rightarrow V^{\otimes d}$ given by $\iota(v_1, v_2, \dots, v_d) = v_1 \otimes v_2 \otimes \dots \otimes v_d$, where $v_i \in V$ for $i = 1, \dots, d$. Any linear map $L : V^{\otimes d} \rightarrow \mathbb{R}$ induces a multi-linear map $\ell = L \circ \iota : V^d \rightarrow \mathbb{R}$. Conversely, every multi-linear map $\ell : V^d \rightarrow \mathbb{R}$ factors as $\ell = L \circ \iota$ for a unique linear map $L : V^{\otimes d} \rightarrow \mathbb{R}$ by the universal property of the tensor product (see [18, Chapter XVI]). This proves the bijection between (2) and (3). Since we have identified V with its dual V^* we can also identify $V^{\otimes d}$ with $(V^*)^{\otimes d} \cong (V^{\otimes d})^*$, which gives the equivalence between (1) and (2). ■

We will frequently switch between these different viewpoints in the lemma.

The group $O(V)$ and the symmetric group Σ_d act on the d -fold tensor product space as follows. Let \mathcal{S} be a rank r tensor in $V^{\otimes d}$ such that $\mathcal{S} = \sum_{i=1}^r v_{1,i} \otimes v_{2,i} \otimes \dots \otimes v_{d,i} \in V^{\otimes d}$, where $v_{j,i} \in V$ for all $i = 1, \dots, r$ and $j = 1, \dots, d$. If $A \in O(V)$, then we have

$$(2.2) \quad A \cdot \mathcal{S} = \sum_{i=1}^r Av_{1,i} \otimes Av_{2,i} \otimes \dots \otimes Av_{d,i}.$$

If $\pi \in \Sigma_d$ is a permutation, then

$$(2.3) \quad \pi \cdot \mathcal{S} = \sum_{i=1}^r v_{\pi^{-1}(1),i} \otimes v_{\pi^{-1}(2),i} \otimes \dots \otimes v_{\pi^{-1}(d),i}.$$

The actions of $O_n(\mathbb{R})$ and Σ_d on $V^{\otimes d}$ commute.

The subspace of $O(V)$ -invariant tensors in $V^{\otimes d}$ is

$$(2.4) \quad (V^{\otimes d})^{O(V)} = \{\mathcal{T} \in V^{\otimes d} : A \cdot \mathcal{T} = \mathcal{T} \text{ for all } A \in O(V)\}.$$

A linear map $L : V^{\otimes d} \rightarrow \mathbb{R}$ is $O(V)$ -invariant if $L(A \cdot \mathcal{T}) = L(\mathcal{T})$ for all tensors \mathcal{T} and all $A \in O(V)$. A multi-linear map $M : V^d \rightarrow \mathbb{R}$ is $O(V)$ -invariant if $M(Av_1, \dots, Av_d) = M(v_1, \dots, v_d)$ for all $v_1, \dots, v_d \in V$ and all $A \in O(V)$.

Corollary 2.2. *There are bijections between the following sets:*

1. $(V^{\otimes d})^{\text{O}(V)}$, the set of $\text{O}(V)$ -invariant tensors in $V^{\otimes d}$;
2. the set of $\text{O}(V)$ -invariant linear maps $V^{\otimes d} \rightarrow \mathbb{R}$;
3. the set of $\text{O}(V)$ -invariant multilinear maps $V^d \rightarrow \mathbb{R}$.

Proof. Following the proof of Lemma 2.1, we see that the bijections in Lemma 2.1 preserve the action of the orthogonal group $\text{O}(V)$ and induce the desired bijections in Corollary 2.2. ■

2.2. The First Fundamental Theorem of Invariant Theory. The First Fundamental Theorem of Invariant Theory for the orthogonal group (Theorem 2.5 below) gives us a description of $(V^{\otimes d})^{\text{O}(V)}$. If d is odd then $(V^{\otimes d})^{\text{O}(V)} = 0$. We now will describe $(V^{\otimes d})^{\text{O}(V)}$ for $d = 2e$ where e is a positive integer.

A **labeled Brauer diagram** of size $d = 2e$ is a perfect matching of a complete graph where the vertices are labeled $1, 2, \dots, d$ (see [8, Chapter 10] for more details).

Example 2.3. Below is a labeled Brauer diagram of size 6:

$$(2.5) \quad \mathbf{D} = \begin{array}{c} 1 \text{ --- } 3 \quad 5 \\ \quad \quad \quad \diagdown \\ 2 \text{ --- } 4 \quad 6 \end{array} .$$

We denote this diagram by $(1\ 3)(2\ 6)(4\ 5)$.

To a labeled Brauer diagram \mathbf{D} of size $d = 2e$ we can associate an $\text{O}(V)$ -invariant multilinear map $\mathcal{M}_{\mathbf{D}} : V^d \rightarrow \mathbb{R}$ as follows. If i_k is connected to j_k for $k = 1, 2, \dots, e$ in the diagram \mathbf{D} , then we define

$$(2.6) \quad \mathcal{M}_{\mathbf{D}}(v_1, v_2, \dots, v_d) = (v_{i_1} \cdot v_{j_1})(v_{i_2} \cdot v_{j_2}) \cdots (v_{i_e} \cdot v_{j_e})$$

for all $v_1, \dots, v_d \in V$. By Corollary 2.2 the $\text{O}(V)$ -invariant multilinear map $\mathcal{M}_{\mathbf{D}}$ corresponds to some $\text{O}(V)$ -invariant linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \rightarrow \mathbb{R}$ and an $\text{O}(V)$ -invariant tensor $\mathcal{T}_{\mathbf{D}} \in (V^{\otimes d})^{\text{O}(V)}$, which we make more explicit now. As in the proof of Lemma 2.1, the universal property of the tensor product gives us a unique linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \rightarrow \mathbb{R}$ such that

$$(2.7) \quad \mathcal{L}_{\mathbf{D}}(v_1 \otimes v_2 \otimes \cdots \otimes v_d) = \mathcal{M}_{\mathbf{D}}(v_1, v_2, \dots, v_d) = (v_{i_1} \cdot v_{j_1})(v_{i_2} \cdot v_{j_2}) \cdots (v_{i_e} \cdot v_{j_e}).$$

By Corollary 2.2, there is also a unique tensor $\mathcal{T}_{\mathbf{D}} \in V^{\otimes d}$ such that $\mathcal{L}_{\mathbf{D}}(A) = \mathcal{T}_{\mathbf{D}} \cdot A$ for all tensors $A \in \text{O}(V)$.

Example 2.4. If \mathbf{D} is the diagram in (2.5), and e_1, \dots, e_n is a basis of V , then we have

$$(2.8) \quad \mathcal{T}_{\mathbf{D}} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n e_i \otimes e_j \otimes e_i \otimes e_k \otimes e_k \otimes e_j.$$

The indices i, j, k correspond to the edges $(1\ 3)$, $(2\ 6)$ and $(4\ 5)$ respectively.

The proof of the following theorem is in Theorem 4.3.3 and Proposition 10.1.3 of [8].

Theorem 2.5 (FFT of Invariant Theory for O_n [8, 24]). *The space $(V^{\otimes d})^{\text{O}(V)}$ of invariant tensors is spanned by all $\mathcal{T}_{\mathbf{D}}$ where \mathbf{D} is a Brauer diagram on d vertices.*

The following result is well-known (see for example [3, 22]), but the idea of the proof is useful later.

Proposition 2.6. *The number of Brauer diagrams (and perfect matchings in a complete graph) for $d = 2e$ vertices is $1 \cdot 3 \cdot 5 \cdots (2e - 1)$.*

Proof. Let N_e be the number of Brauer diagrams on $2e$ nodes. Clearly $N_1 = 1$. We can construct $2e + 1$ Brauer diagrams on $2e + 2$ nodes from a Brauer diagram \mathbf{D} on $2e$ nodes as follows. We take \mathbf{D} and add two nodes, $2e + 1$ and $2e + 2$. First, we can choose an integer k with $1 \leq k \leq 2e$ and let l be the vertex that k is connected to. Then we disconnect k from l , connect $2e + 1$ to k and connect $2e + 2$ to l . This gives us a Brauer diagram \mathbf{D}_k on $2e + 2$ nodes. Alternatively, we can also connect $2e + 1$ to $2e + 2$ and get a Brauer diagram on $2e + 2$ nodes that we call \mathbf{D}_{2e+1} . Thus, we have constructed Brauer diagrams $\mathbf{D}_1, \dots, \mathbf{D}_{2e+1}$ from \mathbf{D} . One can verify that we generate all Brauer diagrams on $2e + 2$ nodes exactly once if we vary \mathbf{D} over all Brauer diagrams on $2e$ nodes. So $N_{e+1} = (2e + 1)N_e$ for all e . ■

2.3. Partial Brauer diagrams. A **partial Brauer diagram** of size d is a graph with d vertices labeled $1, 2, \dots, d$ whose edges form a partial matching. Our convention is to draw loose edges at the vertices that are not matched. To a partial Brauer diagram \mathbf{D} with e edges, we can associate an $O(V)$ -invariant multi-linear map $\mathcal{M}_{\mathbf{D}} : V^d \rightarrow V^{\otimes(d-2e)}$ and a linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \rightarrow V^{\otimes d-2e}$.

Example 2.7. For the diagram:

$$(2.9) \quad \mathbf{D} = \begin{array}{ccccc} & 1 & \text{---} & 3 & & 5 \\ & & & & & / \\ 2 & & & 4 & & 6 \\ | & & & & & | \end{array}$$

we have

$$(2.10) \quad \mathcal{M}_{\mathbf{D}}(v_1, v_2, \dots, v_6) = (v_1 \cdot v_3)(v_4 \cdot v_5)v_2 \otimes v_6 \in V^{\otimes 2}$$

for $v_1, v_2, \dots, v_6 \in V$.

Before giving the general rule of computing inner products of tensors associated to Brauer diagrams, we give an illustrative example.

Example 2.8. We compute the inner product $\mathcal{T}_{\mathbf{D}_1}$ and $\mathcal{T}_{\mathbf{D}_2}$ where \mathbf{D}_1 and \mathbf{D}_2 are the diagrams below:

$$(2.11) \quad \mathbf{D}_1 = \begin{array}{ccccc} & 1 & \text{---} & 3 & & 5 \\ & & & & & / \\ 2 & & & 4 & & 6 \\ \text{---} & & & & & \end{array} \quad \mathbf{D}_2 = \begin{array}{ccccc} & 1 & & 3 & & 5 \\ & & & & & / \\ 2 & & & 4 & & 6 \\ & & & & & \backslash \end{array}.$$

We get

$$(2.12) \quad \mathcal{T}_{\mathbf{D}_1} \cdot \mathcal{T}_{\mathbf{D}_2} = \left(\sum_{i,j,k} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_i \otimes \mathbf{e}_k \otimes \mathbf{e}_k \otimes \mathbf{e}_j \right) \cdot \left(\sum_{p,q,r} \mathbf{e}_p \otimes \mathbf{e}_p \otimes \mathbf{e}_q \otimes \mathbf{e}_r \otimes \mathbf{e}_r \otimes \mathbf{e}_q \right) = \sum_{i,j,k,p,q,r} (\mathbf{e}_i \cdot \mathbf{e}_p)(\mathbf{e}_j \cdot \mathbf{e}_p)(\mathbf{e}_i \cdot \mathbf{e}_q)(\mathbf{e}_k \cdot \mathbf{e}_r)(\mathbf{e}_k \cdot \mathbf{e}_r)(\mathbf{e}_j \cdot \mathbf{e}_q).$$

To get a nonzero summand we have to have $i = j = p = q$ and $k = r$. The result of the summation is $\sum_{i=1}^n \sum_{k=1}^n 1 = n^2$. We can visualize this computation as follows

$$(2.13) \quad \begin{array}{c} \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \end{array} \cdot \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} = \begin{array}{c} \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \\ \bullet \text{---} \bullet \end{array} = n^2.$$

The edges of \mathbf{D}_1 correspond to the indices i, j, k and the edges of \mathbf{D}_2 correspond to the indices p, q, r . We overlay the diagrams. The indices of the edges in a cycle must all be the same. Since there are two cycles, namely i, p, j, q and k, r we essentially sum over two indices and get n^2 .

The general rule is clear now.

Corollary 2.9. *The dot product of two tensors $\mathcal{T}_{\mathbf{D}_1}, \mathcal{T}_{\mathbf{D}_2} \in V^{\otimes d}$ can be computed as follows. We overlay the two diagrams \mathbf{D}_1 and \mathbf{D}_2 so that the (labeled) nodes coincide. Then $\mathcal{T}_{\mathbf{D}_1} \cdot \mathcal{T}_{\mathbf{D}_2} = n^k$ where k is the number of cycles (including 2-cycles).*

Proof. The tensor $\mathcal{T}_{\mathbf{D}_1}$ is equal to the sum of all $\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \cdots \otimes \mathbf{e}_{i_d}$ with $1 \leq i_1, i_2, \dots, i_d \leq n$ such that $i_j = i_k$ whenever (j, k) is an edge in \mathbf{D}_1 . Similarly, $\mathcal{T}_{\mathbf{D}_2}$ is the sum of all $\mathbf{e}_{p_1} \otimes \mathbf{e}_{p_2} \otimes \cdots \otimes \mathbf{e}_{p_d}$ with $1 \leq p_1, p_2, \dots, p_d \leq n$ and $p_j = p_k$ whenever (j, k) is an edge in \mathbf{D}_2 . We compute

$$(2.14) \quad \mathcal{T}_{\mathbf{D}_1} \cdot \mathcal{T}_{\mathbf{D}_2} = \sum_{i_1, \dots, i_d, p_1, \dots, p_d} (\mathbf{e}_{i_1} \cdot \mathbf{e}_{p_1})(\mathbf{e}_{i_2} \cdot \mathbf{e}_{p_2}) \cdots (\mathbf{e}_{i_d} \cdot \mathbf{e}_{p_d}),$$

where the sum is over all $(i_1, \dots, i_d, p_1, \dots, p_d)$ for which $i_j = i_k$ when (j, k) is an edge in \mathbf{D}_1 and $p_j = p_k$ when (j, k) is an edge in \mathbf{D}_2 . The summand $(\mathbf{e}_{i_1} \cdot \mathbf{e}_{p_1})(\mathbf{e}_{i_2} \cdot \mathbf{e}_{p_2}) \cdots (\mathbf{e}_{i_d} \cdot \mathbf{e}_{p_d})$ is equal to 1 if $i_j = p_j$ for all j , and 0 otherwise. Setting i_j equal to p_j corresponds to overlaying the diagrams \mathbf{D}_1 and \mathbf{D}_2 . So (2.14) is equal to the number of tuples (i_1, i_2, \dots, i_d) with $1 \leq i_1, i_2, \dots, i_d \leq n$, and $i_j = i_k$ whenever (j, k) is an edge in \mathbf{D}_1 or in \mathbf{D}_2 . If k is the number of cycles, then we can choose exactly k of indices i_1, i_2, \dots, i_d freely in the set $\{1, 2, \dots, n\}$, and the other indices are uniquely determined by these choices. This proves that (2.14) is equal to n^k . ■

The proof of the following proposition is similar to the proof of Proposition 2.6.

Proposition 2.10. *Suppose that \mathbf{E} is a Brauer diagram on $d = 2e$ vertices, and $\mathcal{S}_d = \sum_{\mathbf{D}} \mathcal{T}_{\mathbf{D}}$, where the sum is over all Brauer diagrams \mathbf{D} on d vertices. Then we have $\mathcal{T}_{\mathbf{E}} \cdot \mathcal{S}_d = n(n+2) \cdots (n+d-2)$.*

Proof. Let \mathbf{E} be a Brauer diagram on $d = 2e$ vertices and let \mathbf{E}' be the diagram obtained from \mathbf{E} by adding the edge $(d+1, d+2)$. Recall that in the proof of Proposition 2.6, we

constructed diagrams $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{d+1}$ from the Brauer diagram \mathbf{D} on d vertices. Note that if $1 \leq i \leq d$, we get $\mathcal{T}_{\mathbf{E}'} \cdot \mathcal{T}_{\mathbf{D}_i} = \mathcal{T}_{\mathbf{E}} \cdot \mathcal{T}_{\mathbf{D}}$ because we get the diagram of $\mathcal{T}_{\mathbf{E}'} \cdot \mathcal{T}_{\mathbf{D}_i}$ from that of $\mathcal{T}_{\mathbf{E}} \cdot \mathcal{T}_{\mathbf{D}}$ by changing one k -cycle to a $(k+2)$ -cycle. We also have $\mathcal{T}_{\mathbf{E}'} \cdot \mathcal{T}_{\mathbf{D}_{d+1}} = n(\mathcal{T}_{\mathbf{E}} \cdot \mathcal{T}_{\mathbf{D}})$ as we are adding one 2-cycle. This shows that $\mathcal{T}_{\mathbf{E}'} \cdot \mathcal{S}_{d+1} = (n+d)\mathcal{T}_{\mathbf{E}} \cdot \mathcal{S}_d$. The proposition then follows by induction and symmetry. \blacksquare

Example 2.11. For $e = 2$, we get

$$(2.15) \quad \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \cdot \left(\begin{array}{|c|} \hline \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \text{---} \bullet \\ \hline \bullet \text{---} \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \text{---} \bullet \\ \hline \bullet \text{---} \bullet \\ \hline \end{array} \right) = \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \begin{array}{|c|} \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \text{---} \bullet \\ \hline \bullet \text{---} \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \text{---} \bullet \\ \hline \bullet \text{---} \bullet \\ \hline \end{array} = n^2 + n + n = n(n+2).$$

2.4. The expected rank 1 unit tensor. Let $S^{n-1} = \{v \in V \mid \|v\| = 1\}$ be the unit sphere equipped with the $O(V)$ -invariant volume form $d\mu$ that is normalized such that $\int_{S^{n-1}} d\mu = 1$.

Proposition 2.12. *If we integrate*

$$(2.16) \quad v^{\otimes 2e} = \underbrace{v \otimes v \otimes \dots \otimes v}_{2e}$$

over S^{n-1} then we get $\int_{S^{n-1}} v^{\otimes 2e} d\mu = \frac{1}{n(n+2)\dots(n+2e-2)} \mathcal{S}_{2e}$.

Proof. Let $U = \int_{S^{n-1}} v^{\otimes 2e} d\mu$. Since U is $O(V)$ -invariant, it is a linear combination of Brauer diagrams. The tensor U is also invariant under the action of the symmetric group Σ_{2e} , where the action of the symmetric group is given in (2.3). This shows that each Brauer diagram appears with the same coefficient in U . So we have $U = C\mathcal{S}_{2e}$ where C is some constant. Let \mathbf{D} be some Brauer diagram on $2e$ vertices. The value of C is obtained from

$$(2.17) \quad 1 = \int_{S^{n-1}} d\mu = \int_{S^{n-1}} (\mathcal{T}_{\mathbf{D}} \cdot v^{\otimes 2e}) d\mu = \mathcal{T}_{\mathbf{D}} \cdot \int_{S^{n-1}} v^{\otimes 2e} d\mu = \\ = C(\mathcal{T}_{\mathbf{D}} \cdot \mathcal{S}_{2e}) = Cn(n+2)\dots(n+2e-2). \quad \blacksquare$$

3. Computations with 3-way tensors.

3.1. Colored Brauer diagrams. We now consider 3 Euclidean \mathbb{R} -vector spaces R, G and B of dimension p, q and r respectively. The tensor product space $V = R \otimes G \otimes B$ is a representation of $H := O(R) \times O(G) \times O(B)$. We keep the notations introduced in Sections 1 and 2 based on the fact that tensor product of vector spaces is a vector space itself. We are interested in H -invariant tensors in $V^{\otimes d}$. We have an explicit linear isomorphism $\psi : R^{\otimes d} \otimes G^{\otimes d} \otimes B^{\otimes d} \rightarrow V^{\otimes d}$ defined by

$$(3.1) \quad \psi((a_1 \otimes \dots \otimes a_d) \otimes (b_1 \otimes \dots \otimes b_d) \otimes (c_1 \otimes \dots \otimes c_d)) = (a_1 \otimes b_1 \otimes c_1) \otimes \dots \otimes (a_d \otimes b_d \otimes c_d),$$

where $a_i \in R, b_i \in G$ and $c_i \in B$ for $i = 1, \dots, d$. Restriction to H -invariant tensors gives an isomorphism

$$(3.2) \quad \psi : (R^{\otimes d})^{O(R)} \otimes (G^{\otimes d})^{O(G)} \otimes (B^{\otimes d})^{O(B)} \rightarrow (V^{\otimes d})^H.$$

It follows from Theorem 2.5 that the space $(R^{\otimes d})^{O(R)}$ of invariant tensors is spanned by tensors corresponding to Brauer diagrams on the set $1, 2, \dots, d$. We will use red edges for

these diagrams. The space $(G^{\otimes d})^{\mathcal{O}(G)}$ is spanned by tensors corresponding to green Brauer diagrams on the set $1, 2, \dots, d$ and the space $(B^{\otimes d})^{\mathcal{O}(B)}$ is spanned by tensors corresponding to blue Brauer diagrams on $1, 2, \dots, d$. Using the isomorphism (3.2), we see that $(V^{\otimes d})^H$ is spanned by diagrams with vertices $1, 2, 3, \dots, d$ and red, green and blue edges such that for each of the colors we have a perfect matching. This means that each vertex has exactly one red, one green and one blue edge.

Definition 3.1. A **colored Brauer diagram** of size $d = 2e$ is a graph with d vertices labeled $1, 2, \dots, d$ and e red, e green and e blue edges such that for each color, the edges of that color form a perfect matching.

A colored Brauer diagram \mathbf{D} on d vertices is an overlay of a red diagram \mathbf{D}_R , a green diagram \mathbf{D}_G and a blue diagram \mathbf{D}_B . To a colored Brauer diagram \mathbf{D} we can associate an invariant tensor $\mathcal{T}_{\mathbf{D}} \in (V^{\otimes d})^H$ by

$$(3.3) \quad \mathcal{T}_{\mathbf{D}} = \psi(\mathcal{T}_{\mathbf{D}_R} \otimes \mathcal{T}_{\mathbf{D}_G} \otimes \mathcal{T}_{\mathbf{D}_B}).$$

Proposition 3.2. The space $(V^{\otimes d})^H$ is spanned by all $\mathcal{T}_{\mathbf{D}}$, where \mathbf{D} is a colored Brauer diagram on d vertices.

Proof. It follows from Theorem 2.5 that the space $(R^{\otimes d})^{\mathcal{O}(R)}$ is spanned by tensors $\mathcal{T}_{\mathbf{D}_R}$ where \mathbf{D}_R is a red Brauer diagram on d vertices. Similarly, $(G^{\otimes d})^{\mathcal{O}(G)}$ is spanned by tensors $\mathcal{T}_{\mathbf{D}_G}$ and $(B^{\otimes d})^{\mathcal{O}(B)}$ is spanned by tensors $\mathcal{T}_{\mathbf{D}_B}$ where \mathbf{D}_G and \mathbf{D}_B are green and blue Brauer diagrams on d vertices respectively. The space $(R^{\otimes d})^{\mathcal{O}(R)} \otimes (G^{\otimes d})^{\mathcal{O}(G)} \otimes (B^{\otimes d})^{\mathcal{O}(B)}$ is spanned by tensors of the form $\mathcal{T}_{\mathbf{D}_R} \otimes \mathcal{T}_{\mathbf{D}_G} \otimes \mathcal{T}_{\mathbf{D}_B}$. Via the isomorphism ψ in (3.2), $(V^{\otimes d})^H$ is spanned by all $\mathcal{T}_{\mathbf{D}} = \psi(\mathcal{T}_{\mathbf{D}_R} \otimes \mathcal{T}_{\mathbf{D}_G} \otimes \mathcal{T}_{\mathbf{D}_B})$ where \mathbf{D} is a colored Brauer diagram that is the overlay of \mathbf{D}_R , \mathbf{D}_G and \mathbf{D}_B . ■

Using the bijections from Lemma 2.1, every colored Brauer diagram \mathbf{D} on d vertices corresponds to a linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \rightarrow \mathbb{R}$ given by $\mathcal{L}_{\mathbf{D}}(\mathcal{A}) = \mathcal{T}_{\mathbf{D}} \cdot \mathcal{A}$ for all $\mathcal{A} \in H$. And this linear map corresponds to a multilinear map $V^d \rightarrow \mathbb{R}$ defined by $\mathcal{M}_{\mathbf{D}}(\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_d) = \mathcal{L}_{\mathbf{D}}(\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \dots \otimes \mathcal{A}_d)$ for all tensors $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_d \in V = R \otimes G \otimes B$.

Corollary 3.3. There are bijections between the following sets:

1. $(V^{\otimes d})^H$;
2. the set of H -invariant linear maps $V^{\otimes d} \rightarrow \mathbb{R}$;
3. the set of H -invariant multilinear maps $V^d \rightarrow \mathbb{R}$.

Proof. The proof is the same as for Corollary 2.2, but with $\mathcal{O}(V)$ replaced by H . ■

For example, the colored Brauer diagram \mathbf{D} :

$$(3.4) \quad \begin{array}{c} 1 \text{ --- } 2 \\ | \quad \times \quad | \\ 3 \text{ --- } 4 \end{array}$$

corresponds to the H -invariant linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes 4} = (R \otimes G \otimes B)^{\otimes 4} \rightarrow \mathbb{R}$ defined by

$$(3.5) \quad (a_1 \otimes b_1 \otimes c_1) \otimes (a_2 \otimes b_2 \otimes c_2) \otimes (a_3 \otimes b_3 \otimes c_3) \otimes (a_4 \otimes b_4 \otimes c_4) \mapsto (a_1 \cdot a_2)(a_3 \cdot a_4)(b_1 \cdot b_4)(b_2 \cdot b_3)(c_1 \cdot c_3)(c_2 \cdot c_4).$$

In a similar way, we define an H -invariant multilinear map $\mathcal{M}_{\mathbf{D}} : V^{\otimes d} \rightarrow \mathbb{R}$ for every colored Brauer diagram \mathbf{D} of size d (i.e., with d vertices). We can view $\mathcal{M}_{\mathbf{D}}$ as a tensor in $(V^{\star})^{\otimes d} \cong V^{\otimes d}$. Viewed as a tensor in $V^{\otimes d}$ we will denote it by $\mathcal{T}_{\mathbf{D}}$.

If \mathbf{D} is a colored Brauer diagram of size d , then the polynomial function defined on V by

$$(3.6) \quad \mathcal{T} \mapsto \mathcal{M}_{\mathbf{D}}(\underbrace{\mathcal{T}, \mathcal{T}, \dots, \mathcal{T}}_d)$$

will be denoted by $\mathcal{P}_{\mathbf{D}}(\mathcal{T})$. The function $\mathcal{P}_{\mathbf{D}}$ is an H -invariant polynomial function on V of degree d . Note that $\mathcal{P}_{\mathbf{D}}$ does not depend on the labeling of the vertices of \mathbf{D} . For example, if we remove the labeling from the diagram \mathbf{D} in (3.4) we get an unlabeled diagram

$$(3.7) \quad \overline{\mathbf{D}} : \begin{array}{c} \square \\ \diagup \quad \diagdown \\ \square \end{array}$$

and we define $\mathcal{P}_{\overline{\mathbf{D}}} = \mathcal{P}_{\mathbf{D}}$. In coordinates, if we write $\mathcal{T} = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r t_{ijk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \in R \otimes G \otimes B$, then we have

$$(3.8) \quad \mathcal{P}_{\mathbf{D}}(\mathcal{T}) = \sum_{a=1}^p \sum_{b=1}^p \sum_{c=1}^q \sum_{d=1}^q \sum_{e=1}^r \sum_{f=1}^r t_{ace} t_{adf} t_{bde} t_{bcf}.$$

Proposition 3.4. *The space of H -invariant polynomial functions on $V = R \otimes G \otimes B$ is spanned by all $\mathcal{P}_{\mathbf{D}}$ where \mathbf{D} is a colored Brauer diagram.*

Proof. Let $(V^{\otimes d})^{\star}$ be the space of multilinear maps, and $\mathbb{R}[V]_d$ be the space of homogeneous polynomial functions on V of degree d . We have a linear map $\gamma : (V^{\otimes d})^{\star} \rightarrow \mathbb{R}[V]_d$ defined as follows: If $\mathcal{M} : V^d \rightarrow \mathbb{R}$ is multilinear, then $\mathcal{P} = \gamma(\mathcal{M})$ is given by

$$(3.9) \quad \mathcal{P}(\mathcal{T}) = \mathcal{M}(\underbrace{\mathcal{T}, \mathcal{T}, \dots, \mathcal{T}}_d).$$

for all $\mathcal{T} \in V$. For a colored Brauer diagram \mathbf{D} we have by definition $\gamma(\mathcal{M}_{\mathbf{D}}) = \mathcal{P}_{\mathbf{D}}$ (see (3.6)). The surjective map γ restricts to a surjective linear map of H -invariant subspaces $((V^{\otimes d})^{\star})^H \rightarrow \mathbb{R}[V]_d^H$, which is also surjective by [8, Lemma 4.2.7]. Since $((V^{\otimes d})^{\star})^H$ is spanned by the tensors $\mathcal{M}_{\mathbf{D}}$ where \mathbf{D} is a colored Brauer diagram, $\mathbb{R}[V]_d^H$ is spanned by all $\mathcal{P}_{\mathbf{D}} = \gamma(\mathcal{M}_{\mathbf{D}})$. ■

If $\mathbf{D} \amalg \mathbf{E}$ is the disjoint union of colored Brauer diagrams, then we have $\mathcal{P}_{\mathbf{D} \amalg \mathbf{E}} = \mathcal{P}_{\mathbf{D}} \mathcal{P}_{\mathbf{E}}$.


Corollary 3.5. *The ring of polynomial H -invariant polynomial functions on V is generated by invariants of the form $\mathcal{P}_{\mathbf{D}}$ where \mathbf{D} is a connected colored Brauer diagram.*

Proof. By Proposition 3.4 the space of H -invariant polynomials is spanned by invariants of the form $\mathcal{P}_{\mathbf{D}}$ where \mathbf{D} is a colored Brauer diagram. We can write $\mathbf{D} = \mathbf{D}_1 \amalg \mathbf{D}_2 \amalg \dots \amalg \mathbf{D}_k$ where \mathbf{D}_i is a connected colored Brauer diagram for every i . We have

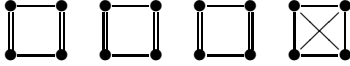
$$(3.10) \quad \mathcal{P}_{\mathbf{D}} = \mathcal{P}_{\mathbf{D}_1} \mathcal{P}_{\mathbf{D}_2} \cdots \mathcal{P}_{\mathbf{D}_k}. \quad \blacksquare$$

Definition 3.6. *We can define $\mathcal{L}_{\mathbf{D}}$, $\mathcal{T}_{\mathbf{D}}$, $\mathcal{M}_{\mathbf{D}}$ and $\mathcal{P}_{\mathbf{D}}$ when \mathbf{D} is a linear combination of diagrams by assuming that these depend linearly on \mathbf{D} . For example, if $\mathbf{D} = \lambda_1 \mathbf{D}_1 + \lambda_2 \mathbf{D}_2 + \dots + \lambda_k \mathbf{D}_k$, then $\mathcal{P}_{\mathbf{D}} = \lambda_1 \mathcal{P}_{\mathbf{D}_1} + \lambda_2 \mathcal{P}_{\mathbf{D}_2} + \dots + \lambda_k \mathcal{P}_{\mathbf{D}_k}$ where $\lambda_i \in \mathbb{R}$ for $i = 1, \dots, k$.*

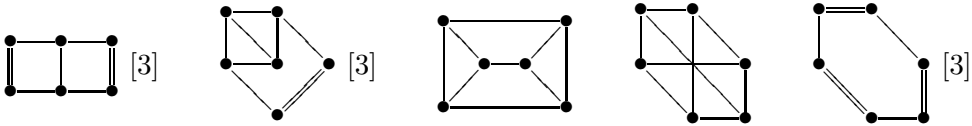
3.2. Generators of polynomial tensor invariants. There is only one connected colored Brauer diagram on 2 vertices:

(3.11) 

There are 4 connected colored (unlabeled) Brauer diagrams on 4 nodes:

(3.12) 

There are 11 connected colored Brauer diagrams on 6 nodes:

(3.13) 

Here, [3] means that by permuting the colors we get 3 pairwise nonisomorphic colored graphs. For $d = 2e$ vertices, the number of connected trivalent colored graphs is given in the following table:

d	2	4	6	8	10	12	14	16	18	20
#	1	4	11	60	318	2806	29359	396196	6231794	112137138





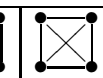
See the Online Encyclopedia of Integer Sequences ([23]), sequence A002831.

Example 3.7. Consider the tensors $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{R}^{n^2 \times n^2 \times n^2}$ given by

(3.14)
$$\mathcal{T}_1 = \frac{1}{n} \sum_{i=1}^{n^2} \mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i \quad \text{and}$$

(3.15)
$$\mathcal{T}_2 = \frac{1}{n\sqrt{n}} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \mathbf{e}_{ni+j+1} \otimes \mathbf{e}_{nj+k+1} \otimes \mathbf{e}_{nk+i+1}.$$

Any flattening of \mathcal{T}_1 and \mathcal{T}_2 is an $n^2 \times n^4$ matrix whose singular values are equal to $\frac{1}{n}$ with multiplicity n^2 . If every vertex in a diagram is adjacent to a double or triple edge, then the corresponding tensor invariant cannot distinguish \mathcal{T}_1 from \mathcal{T}_2 . In the table below, we see that only the tetrahedron diagram can distinguish \mathcal{T}_1 and \mathcal{T}_2 . This invariant captures information from the tensor that cannot be seen in any flattening.

					
\mathcal{T}_1	1	n^{-2}	n^{-2}	n^{-2}	n^{-2}
\mathcal{T}_2	1	n^{-2}	n^{-2}	n^{-2}	1

3.3. Complexity of Tensor Invariants. A polynomial tensor invariant corresponding to a colored Brauer diagram can be computed from subdiagrams.

Example 3.8. To compute

$$(3.16) \quad \begin{array}{c} \bullet & & \bullet \\ | & & | \\ \bullet & & \bullet \\ | & & | \\ \bullet & & \bullet \end{array}$$

we could first compute the partial colored Brauer diagram

$$(3.17) \quad \begin{array}{c} | \\ \bullet \\ | \\ \bullet \\ | \\ \bullet \\ | \end{array} \cdot$$

This partial diagram corresponds to a (symmetric) tensor in $R \otimes R$. If $\mathcal{T} = (t_{ijk})$ then this diagram corresponds to a $p \times p$ matrix $A = (a_{ij})$ where $a_{ij} = \sum_{k=1}^q \sum_{\ell=1}^r t_{ik\ell} t_{j\ell k}$. In practice, one can compute A by first flattening \mathcal{T} to a $p \times (qr)$ matrix B and using $A = BB^t$, where B^t is the transpose of B . The space complexity of this operation is $O(pqr + p^2)$ (we just have to store the tensor \mathcal{T} and the matrix A) and the time complexity is $O(p^2qr)$, because for each pair (i, j) with $1 \leq i, j \leq p$, we have to do $O(qr)$ multiplications and additions. Finally, we compute the invariant as an inner product:

$$(3.18) \quad \begin{array}{c} \bullet & & \bullet \\ | & & | \\ \bullet & & \bullet \\ | & & | \\ \bullet & & \bullet \end{array} = \begin{array}{c} | \\ \bullet \\ | \\ \bullet \\ | \end{array} \cdot \begin{array}{c} | \\ \bullet \\ | \\ \bullet \\ | \end{array}$$

The space complexity of this step is $O(p^2)$ and the time complexity is $O(p^2)$. We conclude that the space complexity of computing (3.16) is $O(pqr + p^2)$ and the time complexity is $O(p^2qr)$. The theoretical time complexity bounds could be improved if we use fast matrix multiplication (such as Strassen's algorithm).

Example 3.9. The invariant

$$(3.19) \quad \begin{array}{c} \bullet & & \bullet \\ | & \diagdown & | \\ \bullet & & \bullet \\ | & \diagup & | \\ \bullet & & \bullet \end{array}$$

is more difficult to compute. We first compute the $p \times p \times q \times q$ tensor \mathcal{U} corresponding to the diagram

$$(3.20) \quad \begin{array}{c} \text{---} & \bullet & \text{---} \\ | & & | \\ \text{---} & \bullet & \text{---} \end{array} \cdot$$

The space complexity of this computation is $O(p^2q^2 + pqr)$ and the time complexity is $O(p^2q^2r)$. Finally we compute

$$(3.21) \quad \begin{array}{c} \bullet & & \bullet \\ | & \diagdown & | \\ \bullet & & \bullet \\ | & \diagup & | \\ \bullet & & \bullet \end{array} = \begin{array}{c} \text{---} & \bullet & \text{---} \\ | & & | \\ \text{---} & \bullet & \text{---} \end{array} \cdot \begin{array}{c} \text{---} & \bullet & \text{---} \\ | & & | \\ \text{---} & \bullet & \text{---} \end{array}$$

An explicit formula for this tensor invariant is given by

$$(3.22) \quad \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^q \sum_{\ell=1}^q \mathcal{U}_{ijk\ell} \mathcal{U}_{ij\ell k}.$$

The space complexity of this step is $O(p^2q^2)$ and the time complexity is $O(p^2q^2)$ as well. Combining the two steps, we see that the time complexity of computing the tetrahedron invariant (3.19) is $O(p^2q^2r)$ and space complexity $O(p^2q^2 + pqr)$. This is the approach we would use if $p \leq q \leq r$. If $p \geq q \geq r$ then a more efficient algorithm is obtained by switching red and blue. In that case we get time $O(pq^2r^2)$.

Example 3.10. To compute

$$(3.23) \quad \begin{array}{c} \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \end{array}$$

we first compute

$$(3.24) \quad \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array}$$

by contracting the $p \times p \times q \times q$ tensor \mathcal{U} with \mathcal{T} . This step requires $O(p^2q^2 + pqr)$ in memory and $O(p^2q^2r)$ in time. From this we compute

$$(3.25) \quad \begin{array}{c} \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \end{array} = \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array} \cdot \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array}.$$

Example 3.11. To compute the below diagram

$$(3.26) \quad \begin{array}{c} \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \end{array}$$

we can first compute

$$(3.27) \quad \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array}$$

which costs $O(p^3qr)$ in memory and $O(p^3q^2r)$ in time, and then we have

$$(3.28) \quad \begin{array}{c} \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \\ \hline \bullet & & \bullet \end{array} = \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array} \cdot \begin{array}{c} \bullet \\ \hline \bullet \\ \hline \bullet \end{array}.$$

It becomes clear that the invariants that correspond to large diagrams can be hard to compute because of memory and time limitations. Some tensor invariants require large tensors in intermediate steps of the computation. There is a method to improve the memory and time requirements with some loss of the accuracy of the result. One can use the Higher Order Singular Value Decomposition (HOSVD) to reduce a $p \times q \times r$ tensor \mathcal{T} to a core tensor \mathcal{T}' of size $p' \times q' \times r'$ where $p' \leq p$, $q' \leq q$ and $r' \leq r$. The HOSVD is a generalization of the singular value decomposition of a matrix, see [19]. If $r > pq$ then the $p \times q \times r$ tensor can be reduced to a $p \times q \times pq$ tensor using HOSVD without any loss at all. HOSVD is a special case of Tucker decomposition [30]. Details of these decomposition methods are beyond the scope of this paper.

4. Approximations of the Spectral Norm.

4.1. The spectral norm. The spectral norm $\|\mathcal{T}\|_\sigma$ of a tensor $\mathcal{T} \in V = R \otimes G \otimes B$ is defined by $\|\mathcal{T}\|_\sigma := \max\{|\mathcal{T} \cdot (x \otimes y \otimes z)| \mid \|x\| = \|y\| = \|z\| = 1\}$. We can view \mathcal{T} as an ℓ_∞ norm on the product of unit spheres $S^{p-1} \times S^{q-1} \times S^{r-1}$. The ℓ_∞ norm is a limit of ℓ_d -norms where $d \rightarrow \infty$. We have $\|\mathcal{T}\|_\sigma := \lim_{d \rightarrow \infty} \|\mathcal{T}\|_{\bar{\sigma},d}$ where

$$(4.1) \quad \|\mathcal{T}\|_{\bar{\sigma},d} := \left(\int_{S^{p-1} \times S^{q-1} \times S^{r-1}} |\mathcal{T} \cdot (x \otimes y \otimes z)|^d d\mu \right)^{1/d}.$$

Suppose that $d = 2e$ is even. We have $|\mathcal{T} \cdot (x \otimes y \otimes z)|^d = \mathcal{T}^{\otimes d} \cdot (x \otimes y \otimes z)^{\otimes d}$ and

$$(4.2) \quad \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} |\mathcal{T} \cdot (x \otimes y \otimes z)|^d d\mu = \mathcal{T}^{\otimes d} \cdot \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes d} d\mu.$$

Up to permutation of the tensor factors, we have the following equality

$$(4.3) \quad \begin{aligned} \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes d} d\mu &= \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x^{\otimes d} \otimes y^{\otimes d} \otimes z^{\otimes d}) d\mu = \\ &= \left(\int_{S^{p-1}} x^{\otimes d} d\mu \right) \otimes \left(\int_{S^{q-1}} y^{\otimes d} d\mu \right) \otimes \left(\int_{S^{r-1}} z^{\otimes d} d\mu \right). \end{aligned}$$

We will normalize the norm $\|\cdot\|_{\bar{\sigma},d}$ so that the value of simple tensors of unit length is equal to 1. So we define a norm $\|\cdot\|_{\sigma,d}$ by

$$\|\mathcal{T}\|_{\sigma,d} = \frac{\|\mathcal{T}\|_{\bar{\sigma},d}}{\|x \otimes y \otimes z\|_{\bar{\sigma},d}},$$

where x, y, z are unit vectors. We have $\lim_{d \rightarrow \infty} \|\mathcal{T}\|_{\sigma,d} = \|\mathcal{T}\|_\sigma$. We will compute $\|\mathcal{T}\|_{\sigma,d}$ for $d = 2$ and $d = 4$.

For any even d , we let $\mathcal{S}_{R,d} \in R^{\otimes d}$ be the sum of all red Brauer diagrams on d vertices. For example,

$$(4.4) \quad \mathcal{S}_{R,4} = \begin{array}{c} \bullet \\ | \\ \bullet \end{array} + \begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array}.$$

Similarly $\mathcal{S}_{G,d}$ and $\mathcal{S}_{B,d}$ are the respective sums of all green and blue Brauer diagrams on d nodes.

4.2. The approximation for $d = 2$.

Proposition 4.1. *The norm $\|\cdot\|_{\sigma,2}$ is equal to the Euclidean (or Frobenius) norm $\|\cdot\|$.*

Proof. If we let $e = 1$, then it follows from Proposition 2.12 that

$$(4.5) \quad \int_{S^{p-1}} x \otimes x d\mu = \frac{1}{p} \mathcal{S}_{R,2}, \quad \int_{S^{q-1}} y \otimes y d\mu = \frac{1}{q} \mathcal{S}_{G,2}, \quad \text{and} \quad \int_{S^{r-1}} z \otimes z d\mu = \frac{1}{r} \mathcal{S}_{B,2}.$$

Therefore, we get

$$(4.6) \quad \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes 2} d\mu = \frac{1}{pqr} \mathcal{S}_{R,2} \otimes \mathcal{S}_{G,2} \otimes \mathcal{S}_{B,2}. \quad \blacksquare$$

In diagrams, we get

$$\int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes 2} d\mu = \frac{1}{pqr} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array}.$$

So we have

$$\|\mathcal{T}\|_{\bar{\sigma},2}^2 = (\mathcal{T} \otimes \mathcal{T}) \cdot \left(\frac{1}{pqr} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \right) = \frac{1}{pqr} \mathcal{T} \cdot \mathcal{T} = \frac{1}{pqr} \|\mathcal{T}\|^2.$$

and $\|\mathcal{T}\|_{\bar{\sigma},2} = \frac{1}{\sqrt{pqr}} \|\mathcal{T}\|$. It follows that $\|\mathcal{T}\|_{\sigma,2}$ is equal to the Euclidean norm $\|\cdot\|$.

4.3. The approximation for $d = 4$.

Theorem 4.2. *We have that $\|\mathcal{T}\|_{\sigma,4}^4 = \mathcal{P}_{\mathbf{D}}(\mathcal{T})$ where*

$$(4.7) \quad \mathbf{D} = \frac{3 \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array}}{27}$$

and $\mathcal{P}_{\mathbf{D}}$ is defined as in (3.8).

Proof. If we employ the Proposition 2.12 for $e = 2$, then we get

$$(4.8) \quad \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes 4} d\mu = \left(\int_{S^{p-1}} x^{\otimes 4} d\mu \right) \otimes \left(\int_{S^{q-1}} y^{\otimes 4} d\mu \right) \otimes \left(\int_{S^{r-1}} z^{\otimes 4} d\mu \right) = \frac{1}{p(p+2)q(q+2)r(r+2)} \mathcal{S}_{R,4} \otimes \mathcal{S}_{G,4} \otimes \mathcal{S}_{B,4}.$$

We calculate

$$(4.9) \quad \mathcal{S}_{R,4} \otimes \mathcal{S}_{G,4} = \left(\begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \right) \otimes \left(\begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \right) =$$

$$\begin{array}{c} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \\ \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \\ \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} + \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array} \end{array} = 3 \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array}.$$

In this calculation, we have omitted the labeling of the vertices. The last equality is only true if we symmetrize the right-hand side over all 24 permutations on the 4 vertices.

$$(4.10) \quad \mathcal{S}_{R,4} \otimes \mathcal{S}_{G,4} \otimes \mathcal{S}_{B,4} = \left(3 \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \parallel & \parallel \\ \bullet & \bullet \end{array} \right) \otimes \left(\begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} \right) =$$

$$3 \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} + 3 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + 3 \begin{array}{c} \bullet & \bullet \\ \parallel & \parallel \\ \bullet & \bullet \end{array} +$$

$$6 \begin{array}{c} \bullet & \bullet \\ \parallel & \parallel \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} = 3 \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \parallel & \parallel \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array}.$$

For this calculation, one should symmetrize the red-green diagrams over all 24 permutations. However, if we do not do this the result will not change because the blue diagrams are symmetrized over all permutations. We conclude that $\|\mathcal{T}\|_{\sigma,4}^4 = \mathcal{P}_{\mathbf{D}}(\mathcal{T})$ where

$$(4.11) \quad \mathbf{D} = \frac{3 \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} \begin{array}{c} \bullet \\ \parallel \\ \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \parallel & \parallel \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} + 6 \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array}}{27}. \quad \blacksquare$$

4.4. Other approximations of the spectral norm. We say that a norm $\|\cdot\|_{\#}$ is a degree d norm if $\|\mathcal{T}\|_{\#}^d$ is a polynomial function on \mathcal{T} of degree d . The norm $\|\cdot\|_{\sigma,d}$ is a norm of degree d . In particular, $\|\cdot\|_{\sigma,4}$ is a norm of degree 4. In this section we study other norms of degree 4 that approximate the spectral norm.

Consider the degree 2 covariant $\mathcal{U} = V \rightarrow G^{\otimes 2} \otimes B^{\otimes 2}$ defined by

$$\mathcal{U} = \mathcal{U}(\mathcal{T}) = \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array}.$$

We have

$$(4.12) \quad 0 \leq \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} \cdot \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} = \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} \cdot \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} = \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} \quad \text{and}$$

$$(4.13) \quad \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} \cdot \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} = \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array}$$

(in this calculation, the diagrams represent their evaluations on $\mathcal{T} \otimes \mathcal{T} \otimes \mathcal{T} \otimes \mathcal{T}$). So we get

$$(4.14) \quad \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} = \frac{1}{2}(\mathcal{U} \cdot \mathcal{U}) \geq 0.$$

Permuting the colors also gives

$$(4.15) \quad \begin{array}{c} \bullet & \bullet \\ \text{---} & \text{---} \\ \bullet & \bullet \end{array} + \begin{array}{c} \bullet & \bullet \\ \diagdown & \diagup \\ \bullet & \bullet \end{array} \geq 0.$$

It follows from (4.12) that

$$(4.16) \quad \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} \geq 0.$$

Adding (4.14), (4.15) and (4.16) gives

$$(4.17) \quad \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + 2 \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} \geq 0.$$

Definition 4.3. We define

$$(4.18) \quad \|\mathcal{T}\|_{\#} = \frac{1}{5^{1/4}} \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} + 2 \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right)^{1/4}.$$

We will show that $\|\mathcal{T}\|_{\#}$ is a norm.

Lemma 4.4. Suppose that $f(x) = f(x_1, x_2, \dots, x_m) \in \mathbb{R}[x_1, \dots, x_m]$ is a homogeneous polynomial of degree $d > 0$ with $f(x) > 0$ for all nonzero $x \in \mathbb{R}^m$ and the Hessian matrix $(\frac{\partial^2 f}{\partial x_i \partial x_j})$ is positive semi-definite. Then $\|x\|_{\#} := f(x)^{1/d}$ is a norm on \mathbb{R}^m .

Proof. It is clear that $\|x\|_{\#} = 0$ if and only if $x = 0$. We have $f(\lambda x) = \lambda^d f(x)$ which implies that d must be even. We get $f(\lambda x)^{1/d} = (\lambda^d f(x))^{1/d} = |\lambda| f(x)^{1/d}$. Because the Hessian is positive semi-definite, the function $f(x)$ is convex and the set $B = \{x \mid f(x) \leq 1\}$ is convex, which is also the unit ball for $\|x\|_{\#}$.

If $x, y \in \mathbb{R}^n$ are nonzero, then we have $\frac{x}{\|x\|_{\#}}, \frac{y}{\|y\|_{\#}} \in B$ and therefore

$$\frac{x+y}{\|x\|_{\#} + \|y\|_{\#}} = \frac{\|x\|_{\#}}{\|x\|_{\#} + \|y\|_{\#}} \cdot \frac{x}{\|x\|_{\#}} + \frac{\|y\|_{\#}}{\|x\|_{\#} + \|y\|_{\#}} \cdot \frac{y}{\|y\|_{\#}} \in B.$$

So

$$\left\| \frac{x+y}{\|x\|_{\#} + \|y\|_{\#}} \right\|_{\#} = \frac{\|x+y\|_{\#}}{\|x\|_{\#} + \|y\|_{\#}} \leq 1.$$

This proves the triangle inequality. ■

Proposition 4.5. The function $\|\cdot\|_{\#}$ is a norm.

Proof. From (4.17) it follows that $\|\cdot\|_{\#}$ is nonnegative. If $\|\mathcal{T}\|_{\#} = 0$ for some tensor, then we have equality in (4.17), (4.16) and (4.12). This implies that

$$\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \end{array} = 0.$$

If A is a $p \times qr$ flattening of \mathcal{T} , then we have $A^t A = 0$ where A^t is the transpose of A . It follows that $A = 0$ and $\mathcal{T} = 0$. To show that $\|\cdot\|_{\#}$ satisfies the triangle inequality, we have to show that the Hessian of $h = \|\cdot\|_{\#}^4$ is nonnegative.

Up to a constant, h is equal to (4.17). We can write

$$h(\mathcal{T} + \mathcal{E}) = h_0(\mathcal{T}, \mathcal{E}) + h_1(\mathcal{T}, \mathcal{E}) + h_2(\mathcal{T}, \mathcal{E}) + h_3(\mathcal{T}, \mathcal{E}) + h_4(\mathcal{T}, \mathcal{E})$$

where $h_i(\mathcal{T}, \mathcal{E})$ is a polynomial function of degree $4-i$ in \mathcal{T} and degree i in \mathcal{E} . Here $h_0(\mathcal{T}, \mathcal{E}) = \|\mathcal{T}\|_{\#}^4$ and $h_4(\mathcal{T}, \mathcal{E}) = \|\mathcal{E}\|_{\#}^4$. The function $h_1(\mathcal{T}, \mathcal{E})$ is linear in \mathcal{E} and this linear function is the gradient at \mathcal{T} . The function $h_2(\mathcal{T}, \mathcal{E})$ is quadratic function in \mathcal{E} and is, up to a constant, the Hessian of h at \mathcal{T} . So we have to show that $h_2(\mathcal{T}, \mathcal{E}) \geq 0$ for all tensors \mathcal{T} and \mathcal{E} .

Let us write a black vertex for the tensor \mathcal{T} and a white vertex for the tensor \mathcal{E} . We get the Hessian of a function in \mathcal{T} by summing all the possible ways of replacing two black vertices by two white vertices. The Hessian of the left-hand side of (4.17) is

$$(4.19) \quad \frac{1}{2}h_2(\mathcal{T}, \mathcal{E}) = \begin{array}{c} \begin{array}{cccc} \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \bullet & \bullet \\ \hline \end{array} & + & \begin{array}{|c|c|} \hline \circ & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} & + & \begin{array}{|c|c|} \hline \bullet & \circ \\ \hline \bullet & \bullet \\ \hline \end{array} & + & 2 \begin{array}{|c|c|} \hline \circ & \circ \\ \hline \bullet & \bullet \\ \hline \end{array} & + \\ \begin{array}{|c|c|} \hline \circ & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} & + & \begin{array}{|c|c|} \hline \bullet & \circ \\ \hline \bullet & \bullet \\ \hline \end{array} & + & \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \circ & \bullet \\ \hline \end{array} & + & 2 \begin{array}{|c|c|} \hline \circ & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} & + \\ \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \circ \\ \hline \end{array} & + & \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \circ \\ \hline \end{array} & + & 2 \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \circ \\ \hline \end{array} & + & 2 \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \circ \\ \hline \end{array} \end{array} \end{array}$$

Let $\mathcal{W} : V \otimes V \rightarrow G^{\otimes 2} \otimes B^{\otimes 2}$ be defined by

$$\mathcal{W}(\mathcal{T}, \mathcal{E}) = \begin{array}{c} \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \circ & \circ \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \circ \\ \hline \bullet & \bullet \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \end{array}.$$

We compute

$$(4.20) \quad 0 \leq \frac{1}{4}(\mathcal{W} \cdot \mathcal{W}) = \begin{array}{c} \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \circ & \circ \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \circ \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} + \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \end{array}.$$

and we have

$$(4.21) \quad 0 \leq \begin{array}{c} \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \cdot \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} = \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline \bullet & \bullet \\ \hline \end{array} \end{array}.$$

Adding (4.20) and (4.21) and all expressions obtained by cyclically permuting the colors red, green and blue yields (4.19). This proves that $h_2(\mathcal{T}, \mathcal{E}) \geq 0$ and completes the proof that $\|\cdot\|_{\#}$ is a norm. ■

Definition 4.6. A spectral-like norm is a norm $\|\cdot\|_X$ in $\mathbb{R}^{p \times q \times r}$ with the following properties:

1. $\|\mathcal{T}\|_X = 1$ if \mathcal{T} is a rank 1 tensor with $\|\mathcal{T}\|_2 = 1$;
2. $\|\mathcal{T}\|_X < 1$ if \mathcal{T} is a tensor of rank > 1 with $\|\mathcal{T}\|_2 = 1$.

Examples of spectral-like norms are the spectral norm $\|\cdot\|_{\sigma}$, the norms $\|\cdot\|_{\sigma, d}$ for $d = 2, 4, \dots$ and $\|\cdot\|_{\#}$.

Definition 4.7. A nuclear-like norm is a norm $\|\cdot\|_Y$ in $\mathbb{R}^{p \times q \times r}$ with the following properties:

1. $\|\mathcal{T}\|_Y = 1$ if \mathcal{T} is a rank 1 tensor with $\|\mathcal{T}\|_2 = 1$;
2. $\|\mathcal{T}\|_Y > 1$ if \mathcal{T} is a tensor of rank > 1 with $\|\mathcal{T}\|_2 = 1$.

A norm $\|\cdot\|_Y$ is the dual of another norm $\|\cdot\|_X$ if

$$\|\mathcal{S}\|_Y = \max\{\mathcal{S} \cdot \mathcal{T} : \|\mathcal{T}\|_X \leq 1\}.$$

A norm $\|\cdot\|_Y$ is the dual of $\|\cdot\|_X$ if and only if $\|\cdot\|_X$ is the dual of $\|\cdot\|_Y$. The dual of a spectral-like norm is a nuclear-like norm.

We are particularly interested in norms that are powerful in distinguishing low rank tensors from high rank tensors. Spectral-like norms are normalized such that rank 1 tensors of unit Euclidean length have norm 1. A possible measure for the rank discriminating power of a spectral-like norm $\|\cdot\|_X$ is the expected value of $\mathbb{E}(\|\mathcal{T}\|_X)$ where $\mathcal{T} \in S^{pqr-1}$ is a random unit tensor in $R \otimes G \otimes B$ (with the uniform distribution over the sphere). A smaller value of $\mathbb{E}(\|\mathcal{T}\|_X)$ means more discriminating power, which is better. In this sense, the spectral norm is the best norm, because for spectral-like norms $\|\cdot\|_X$ we have $\|\mathcal{T}\|_X \geq \|\mathcal{T}\|_\sigma$, so $\mathbb{E}(\|\mathcal{T}\|_X) \geq \mathbb{E}(\|\mathcal{T}\|_\sigma)$. We may not be able to compute the value $\mathbb{E}(\|\mathcal{T}\|_X)$ for many norms $\|\cdot\|_X$. If we fix the size of the tensor we can estimate $\mathbb{E}(\|\mathcal{T}\|_X)$ by taking the average of random unit vectors x .

We will compare the norms $\|\cdot\|_{\sigma,4}$ and $\|\cdot\|_{\#}$, which both have degree 4. Although we are not able to give closed formulas for $\mathbb{E}(\|\mathcal{T}\|_{\sigma,4})$ and $\mathbb{E}(\|\mathcal{T}\|_{\#})$, we can compute $\mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4)$ and $\mathbb{E}(\|\mathcal{T}\|_{\#}^4)$. First we note that

$$(4.22) \quad \mathbb{E}(\mathcal{T} \otimes \mathcal{T} \otimes \mathcal{T} \otimes \mathcal{T}) = \frac{1}{pqr(pqr+2)} \left(\begin{array}{c} \bullet \\ | \\ \bullet \end{array} \begin{array}{c} \bullet \\ | \\ \bullet \end{array} + \begin{array}{c} \bullet \text{---} \bullet \\ | \text{---} | \\ \bullet \text{---} \bullet \end{array} + \begin{array}{c} \bullet \text{---} \bullet \\ / \text{---} \backslash \\ \bullet \text{---} \bullet \end{array} \right),$$

because we can view \mathcal{T} as a random unit tensor in $V \otimes V \otimes V \otimes V$ and apply Proposition 2.12.

To compute $\mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4)$ we compute the inner product between (4.22) and (4.7). To perform this computation we overlay the two diagrams and count the number of cycles for each color.

$$\begin{array}{c} \bullet \text{---} \bullet \\ | \text{---} | \\ \bullet \text{---} \bullet \end{array} \cdot \begin{array}{c} \bullet \\ | \\ \bullet \end{array} \begin{array}{c} \bullet \\ | \\ \bullet \end{array} = \begin{array}{c} \bullet \text{---} \bullet \\ | \text{---} | \\ \bullet \text{---} \bullet \end{array} \begin{array}{c} \bullet \\ | \\ \bullet \end{array} \begin{array}{c} \bullet \\ | \\ \bullet \end{array} = pq^2r^2.$$

The result is

$$(4.23) \quad \mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4) = \frac{1}{9pqr(pqr+2)} ((p^2q^2r^2 + 2pqr) + 2(pq^2r^2 + p^2qr + pqr) + 2(p^2qr^2 + pq^2r + pqr) + 2(p^2q^2r + pqr^2 + pqr) + 2(p^2qr + pq^2r + pqr^2)) = \frac{pqr + 2(pq + pr + qr) + 4(p + q + r) + 8}{9(pqr + 2)}.$$

A similar computation shows that

$$(4.24) \quad \mathbb{E}(\|\mathcal{T}\|_{\#}^4) = \frac{(pq + pr + qr) + 3(p + q + r) + 3}{5(pqr + 2)}.$$

The following proposition shows that, in some sense, $\|\cdot\|_{\#}$ is better than $\|\cdot\|_{\sigma,4}$ as an approximation to the spectral norm $\|\cdot\|_\sigma$.

Proposition 4.8. *If $p, q, r \geq 1$ then we have $\mathbb{E}(\|\mathcal{T}\|_\sigma^4) \leq \mathbb{E}(\|\mathcal{T}\|_{\#}^4) \leq \mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4)$ for a random tensor \mathcal{T} sampled from the uniform distribution on the unit sphere. The inequality is strict when two of the numbers p, q, r are at least 2.*

Proof. We calculate

$$\begin{aligned}
(4.25) \quad \mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4) - \mathbb{E}(\|\mathcal{T}\|_{\#}^4) &= \\
&= \frac{pqr + 2(pq + pr + qr) + 4(p + q + r) + 8}{9(pqr + 2)} - \frac{(pq + pr + qr) + 3(p + q + r) + 3}{5(pqr + 2)} = \\
&= \frac{5pqr + (pq + qr + rp) - 7(p + q + r) + 13}{45(pqr + 2)} = \\
&= \frac{5(p-1)(q-1)(r-1) + 6((p-1)(q-1) + (p-1)(r-1) + (q-1)(r-1))}{45(pqr + 2)} \quad \blacksquare
\end{aligned}$$

Remark 4.9. If $p = q = r = n$, then asymptotically, we have that $\mathbb{E}(\|\mathcal{T}\|_{\sigma,4}^4) = O(1)$ and $\mathbb{E}(\|\mathcal{T}\|_{\#}^4) = O(\frac{1}{n})$.

5. Low rank amplification. As motivation, we will first consider a map from matrices to matrices that enhances the low rank structure.

5.1. Matrix amplification. Suppose A is a matrix with singular values $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$. Then we can write $A = U\Sigma V^*$ where U, V are orthogonal, Σ is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_r$, and V^* is the conjugate transpose of V . We have

$$(5.1) \quad AA^*A = (U\Sigma V^*)(V\Sigma^*U^*)(U\Sigma V^*) = U\Sigma^3V^*.$$

The matrix AA^*A has singular values $\lambda_1^3 \geq \lambda_2^3 \geq \dots \geq \lambda_r^3 \geq 0$. If $\lambda_1 > \lambda_2$ then the ratio of the two largest singular values increases from λ_1/λ_2 to λ_1^3/λ_2^3 . If we define a map θ by

$$(5.2) \quad \theta(A) = \frac{AA^*A}{\|AA^*A\|},$$

where $\|\cdot\|$ is the Euclidean (Frobenius) norm, then $\lim_{n \rightarrow \infty} \theta^n(A)$ will converge to a rank 1 matrix $B = UDV^*$ where

$$(5.3) \quad D = \begin{pmatrix} 1 & 0 & \dots \\ 0 & 0 & \\ \vdots & & \ddots \end{pmatrix}.$$

Note that the convergence is very fast. After n iterations of θ , the ratio of the two largest singular values is $(\frac{\lambda_1}{\lambda_2})^{3^n}$. We have that $A \cdot B = \Sigma \cdot D = \lambda_1$ is the spectral norm of A and $\lambda_1 B$ is the best rank 1 approximation of A in the following sense: if C is a rank 1 matrix such that $\|A - C\|$ is minimal, then $C = \lambda_1 B$.

The map θ increases the highest singular value relative to the other singular values. In this sense, θ amplifies the sparse structure of the matrix (meaning low rank in this context).

The map θ is related to the 4-Schatten norm, defined by $\|A\|_{s,4} = \text{trace}((AA^*)^2)^{\frac{1}{4}}$. Namely, the gradient of the function $\|A\|_{s,4}^4$ is $4AA^*A$ and the gradient of the function $\|A\|_{s,4}$ is equal to AA^*A up to a scalar function.

5.2. Tensor amplification. We will now consider amplification of the low rank structure of tensors. For this we take the gradient of a spectral-like norm. Let $h = \|\cdot\|_{\#}^4$. As before, we can write $h(\mathcal{T} + \mathcal{E}) = h_0(\mathcal{T}, \mathcal{E}) + h_1(\mathcal{T}, \mathcal{E}) + h_2(\mathcal{T}, \mathcal{E}) + h_3(\mathcal{T}, \mathcal{E}) + h_4(\mathcal{T}, \mathcal{E})$, where h_i has degree i in \mathcal{E} and degree $4 - i$ in \mathcal{T} . Now $h_0(\mathcal{T}, \mathcal{E}) = \|\mathcal{T}\|_{\#}^4$, the function $\mathcal{E} \mapsto h_1(\mathcal{T}, \mathcal{E})$ is the gradient of h at \mathcal{T} , and $h_2(\mathcal{T}, \mathcal{E})$ is the Hessian that we have already computed. To find a formula for the gradient $h_1(\mathcal{T}, \mathcal{E})$ we express $h(\mathcal{T})$ in diagrams and replace each diagram by all diagrams obtained by replacing one of the closed vertices by an open vertex. Using (4.18) we get

$$(5.4) \quad h(\mathcal{T}) = \|\mathcal{T}\|_{\#}^4 = \frac{1}{5} \left(\begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} + \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} + \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} \right).$$

The gradient is now equal to

$$(5.5) \quad (\nabla h)(\mathcal{T}) = \frac{4}{5} \left(\begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} \right).$$

We can also view these diagrams with an open vertex as partial colored Brauer diagrams by removing the open vertex. For example,

$$(5.6) \quad \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} = \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \begin{array}{c} \bullet \\ \text{---} \\ \bullet \end{array}.$$

Let $\Phi_{\#}(\mathcal{T}) = (\nabla h)(\mathcal{T})$. We view $\Phi_{\#}$ as a polynomial map from $V = R \otimes G \otimes B$ to itself. This map enhances the low rank structure of a tensor \mathcal{T} .

In a similar fashion, we can associate an amplification map $\Phi_{\sigma,4}$ to the norm $\|\cdot\|_{\sigma,4}$. Using (4.7) and similar calculations as before, we get

$$(5.7) \quad \Phi_{\sigma,4}(\mathcal{T}) = \frac{4}{9} \left(\begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \bullet \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} + 2 \begin{array}{c} \bullet \\ \text{---} \\ \bullet \\ \text{---} \\ \circ \end{array} \right).$$

5.3. Tensor amplification and Alternating Least Squares. As we discussed in Section 1.1, Alternating Least Squares (ALS) is a standard approach to find low rank approximations of tensors. For rank 1, this algorithm is particularly simple. For a tensor $\mathcal{T} \in \mathbb{R}^{p \times q \times r}$ we try to find a rank one tensor $a \otimes b \otimes c$ such that $\|\mathcal{T} - a \otimes b \otimes c\|$ is minimal. Here $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$ and $c \in \mathbb{R}^r$. Unlike for higher rank, a best rank 1 approximation always exists. The Alternating Least Squares algorithm works as follows. We start with an initial guess $a \otimes b \otimes c$. Then we fix b and c and update the vector a such that $\|\mathcal{T} - a \otimes b \otimes c\|$ is minimal. This is a least squares regression problem that is easy to solve. Next, we fix a and c and update b , and then we fix a and b and update c . We repeat the process of updating a, b, c until the desired level of convergence is achieved. Numerical experiments were performed using the software MATLAB, along with the `cp_als` implementation of the ALS algorithm from the package `Tensor Toolbox` ([1]). ALS is sensitive to the choice of the initial guess.

The default initialization for `cp_als` is to use a random initial guess. We will also consider a method that we call the *Quick Rank 1* method. For a matrix it is easy to find the best rank 1 approximation from the singular value decomposition. If a real matrix M has a singular

value decomposition $M = \sum_{i=1}^s \lambda_i a_i b_i^T$ where $a_1, \dots, a_s, b_1, \dots, b_s$ are unit orthogonal vectors and $\lambda_1 \geq \lambda_2 \geq \dots \lambda_s \geq 0$ are real numbers, then a best rank 1 approximation of M is $\lambda_1 a_1 b_1^T$ and $M \cdot a_1 b_1^T = \lambda_1$ is the spectral norm of M . (The best rank 1 approximation is unique when $\lambda_1 > \lambda_2$.)

For the *Quick Rank 1* method, we use the matrix case to find an initial rank 1 approximation of a given tensor \mathcal{T} of size $p \times q \times r$. We flatten (unfold) this tensor to a $p \times qr$ matrix $\mathcal{T}_{(1)}$. Then we find a rank 1 approximation of this matrix as described above. Let $\mathcal{T}_{(1)} \approx \beta a d^T$ where a, d are orthogonal vectors and $\beta \in \mathbb{R}$. We convert the vector d of dimension qr to a $q \times r$ matrix D . Now we find the best rank 1 approximation of $D \approx \gamma b c^T$ such that b, c are orthogonal vectors and $\gamma \in \mathbb{R}$. We will use $(\beta\gamma)a \otimes b \otimes c$ as a rank 1 approximation to the tensor \mathcal{T} .

Tensor amplification can be used to obtain better initial guesses for ALS, so that better rank 1 approximations can be found using fewer iterations in the ALS algorithm. We will consider 4 different ways of choosing an initial guess for ALS:

1. **Random.** We choose a random initial guess for the rank 1 approximation.
2. **Quick Rank 1.** We first use the quick rank 1 method described above.
3. **$\Phi_{\sigma,4}$ and Quick Rank 1.** We apply the Quick Rank 1 method to $\Phi_{\sigma,4}(\mathcal{T})$.
4. **$\Phi_{\#}$ and Quick Rank 1.** We apply the Quick Rank 1 method to $\Phi_{\#}(\mathcal{T})$.

Rank 1 approximation methods given above can be generalized to higher ranks. Low rank tensor approximation problem is given in (1.4) and (1.5). Let $\mathcal{T} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ be a tensor of order 3. We will look for a rank $r \geq 2$ approximation \mathcal{S} such that

$$(5.8) \quad \|\mathcal{T} - \mathcal{S}\| \text{ is minimal with } \mathcal{S} = \llbracket \Lambda ; U^{(1)}, U^{(2)}, U^{(3)} \rrbracket$$

where the factor matrices $U^{(i)} \in \mathbb{R}^{p_i \times r}$ for $1 \leq i \leq 3$ and $\Lambda \in \mathbb{R}^r$.

ALS method starts with a random initial guess for the factor matrices. We first fix $U^{(2)}$ and $U^{(3)}$ to solve for $U^{(1)}$, then fix $U^{(1)}$ and $U^{(3)}$ to solve for $U^{(2)}$, and then fix $U^{(1)}$ and $U^{(2)}$ to solve for $U^{(3)}$. This iterative process continues until some convergence criterion is satisfied.

For the *iterative Quick Rank 1 method*, we first employ the Quick Rank 1 method to approximate \mathcal{T} with a rank 1 tensor $\lambda_1 a_1 \otimes b_1 \otimes c_1$. The process continues iteratively; and at each step Quick Rank 1 method is used to find a rank 1 approximation of $\mathcal{T} - \sum_{i=1}^s \lambda_i a_i \otimes b_i \otimes c_i$ for $2 \leq s \leq r - 1$.

As in the rank 1 case, we use 4 different methods to choose an initial guess for the ALS method for the low rank r decomposition of \mathcal{T} :

1. **Random.** We choose a random initial guess for the factor matrices.
2. **Quick Rank 1.** We use an iterative approach based on Quick Rank 1 method as described above. (Algorithm 5.1, $k = 0$.)
3. **$\Phi_{\sigma,4}$ and Quick Rank 1.** We iteratively apply the Quick Rank 1 method to $\Phi_{\sigma,4}(\mathcal{T})$. (Algorithm 5.1, $k = 1$.)
4. **$\Phi_{\#}$ and Quick Rank 1.** We iteratively apply the Quick Rank 1 method to $\Phi_{\#}(\mathcal{T})$. (Algorithm 5.1, $k = 2$.)

6. Experiments.

6.1. Rank 1 approximation. In our experiments, we started with a random $30 \times 30 \times 30$ unit tensor of rank 1, $\mathcal{T} = a \otimes b \otimes c$, where $a, b, c \in \mathbb{R}^{30}$ are random unit vectors, independently

Algorithm 5.1 Low Rank approximation to tensor \mathcal{T} of order 3

```
1: function rank_r_methods( $\mathcal{T}, r, k$ )
2:    $\mathcal{D} = \mathcal{T}$ 
3:    $s = 0$ 
4:   while  $s < r$  do
5:      $s \leftarrow s + 1$ 
6:     if  $k = 0$  then
7:        $\mathcal{U} \leftarrow \mathcal{D}$ 
8:     else if  $k = 1$  then
9:        $\mathcal{U} \leftarrow \Phi_{\sigma,4}(\mathcal{D})$ 
10:    else
11:       $\mathcal{U} \leftarrow \Phi_{\#}(\mathcal{D})$ 
12:      Approximate  $\mathcal{U}$  with a unit rank 1 tensor via Quick rank 1 method:  $\mathcal{U} \approx \lambda_s v_s =$ 
        $\lambda_s a_s \otimes b_s \otimes c_s$ 
13:      Update the coefficients  $\lambda_1, \dots, \lambda_s$  such that  $\|\mathcal{D}\|$  is minimal, where  $\mathcal{D} = \mathcal{T} -$ 
        $\sum_{i=1}^s \lambda_i v_i$ 
14:      return decomposition  $\mathcal{T} = \mathcal{S} + \mathcal{D}$  where  $\mathcal{S} = \sum_{i=1}^r \lambda_i v_i$ 
15:
```

drawn from the uniform distribution on the unit sphere. We then added a random tensor \mathcal{E} of size $30 \times 30 \times 30$ with $\|\mathcal{E}\| = 10$ to obtain a noisy tensor $\mathcal{T}_n = \mathcal{T} + \mathcal{E}$. The noise tensor \mathcal{E} is chosen from the sphere in $\mathbb{R}^{30 \times 30 \times 30} \cong \mathbb{R}^{27000}$ of radius 10 with uniform distribution. Note that there is more noise than the original signal. The signal to noise ratio is $20 \log_{10}(1/10) = -20$ dB. We used four methods for rank 1 approximation. Each method gives a rank 1 tensor $\lambda a' \otimes b' \otimes c'$. To measure how good the rank 1 approximation is to the original tensor \mathcal{T} , we compute the inner product

$$\mathcal{T} \cdot (a' \otimes b' \otimes c') = (a \otimes b \otimes c) \cdot (a' \otimes b' \otimes c') = (a \cdot a')(b \cdot b')(c \cdot c'),$$

which we will call the fit. The fit is a number between 0 and 1 where 1 means a perfect fit.

We created 1000 noisy tensors of size $30 \times 30 \times 30$ as described above. We ran each of the 4 methods to find the best rank 1 approximation for each of the 1000 tensors. For the random initial guess method, we repeated the calculation 10 times with different random initial guesses and recorded the best fit, total number of ALS iterations, and total running time. All other methods were only run once and the fit, total number of ALS iterations, and running time were calculated. For all records, we took the average and standard deviation.

There is a tolerance parameter ε in the ALS implementation in `Tensor Toolbox`. The algorithm terminates if the fit after an iteration increases by a factor smaller than $1 + \varepsilon$. For the default value $\varepsilon = 10^{-4}$ we obtained the following results:

It can be observed from Table 6.1 that a better fit is obtained by using tensor amplification rather than a random initial guess. Even if we take the best case of repeating ALS for 10 different random initial conditions, quick rank with amplification still yields a better fit. The total number of ALS iterations with random initial guess is much larger than for the quick

Random (10 runs)	Max Fit	Total # Iterations	Total Time
Average	0.7136	77.5080	0.0943
Standard Deviation	0.2715	12.0254	0.0159
Quick Rank 1	Fit	# Iterations	Time
Average	0.7848	2.94	0.0177
Standard Deviation	0.1618	1.2345	0.0025
$\Phi_{\sigma,4}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.8010	2	0.0210
Standard Deviation	0.1256	0	0.0027
$\Phi_{\#}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.8178	2	0.0205
Standard Deviation	0.0515	0	0.0025

Table 6.1

A comparison of rank-1 approximation methods with tolerance parameter $\varepsilon = 10^{-4}$

rank 1 initialization, or quick rank 1 with tensor amplification. On average, the number of iterations for the best run with random initialization is 10.44, which is much larger than the number of iterations after tensor amplification, which is 2. The running time is also favorable for the quick rank 1 initialization. Amplification gives a better fit for the rank 1 approximation, while the running time has only marginally increased.

If we change the tolerance to $\varepsilon = 10^{-6}$ then the number of iterations increases and the results are given in Table 6.2. As shown in the table, the amplification $\Phi_{\#}$ performs better than the amplification $\Phi_{\sigma,4}$. This is expected, as the norm $\|\cdot\|_{\#}$ is a better approximation for the spectral norm than $\Phi_{\sigma,4}$. We see that the amplification $\Phi_{\#}$ combined with the quick rank 1 method still yields a better fit than the best-out-of-10 runs with random initialization. The number of iterations for the random initialization approximation with the best fit is 25.78 on average, while the average number of ALS iterations for $\Phi_{\#}$ combined with quick rank 1 is only 3.54.

6.2. Rank 2 approximation. We started with a random $40 \times 40 \times 40$ unit tensor of rank 2, $\mathcal{T} = a_1 \otimes b_1 \otimes c_1 + a_2 \otimes b_2 \otimes c_2$, where $a_1, b_1, c_1, a_2, b_2, c_2 \in \mathbb{R}^{40}$ are random unit vectors, independently drawn from the uniform distribution on the unit sphere. We then added a random tensor \mathcal{E} of size $40 \times 40 \times 40$ with $\|\mathcal{E}\| = 10$ to obtain a noisy tensor $\mathcal{T}_n = \mathcal{T} + \mathcal{E}$. The noise tensor \mathcal{E} is chosen from the sphere in $\mathbb{R}^{40 \times 40 \times 40} \cong \mathbb{R}^{64000}$ of radius 10 with uniform distribution. Each method gives a rank 2 tensor \mathcal{S} of size $40 \times 40 \times 40$ and the fit of the approximation is given by $(\mathcal{T} \cdot \mathcal{S})/\|\mathcal{S}\|$. As in Section 6.1, we created 1000 noisy tensors of size $40 \times 40 \times 40$ and we ran each of the 4 methods to find a best rank 2 approximation for each tensor. Random initial guess method is repeated 10 times for each tensor and the best fits, total number of iterations and total running times were recorded. The other three methods were run only once and the fit, total number of ALS iterations, and running time were recorded for each tensor. For the tolerance parameter $\varepsilon = 10^{-4}$, the average and the standard deviation of all the records are given in Table 6.3.

Random (10 runs)	Max Fit	Total # Iterations	Total Time
Average	0.8120	290.3230	0.2893
Standard Deviation	0.0914	82.7586	0.0803
Quick rank 1	Fit	# Iterations	Time
Average	0.7955	6.9780	0.0210
Standard Deviation	0.1436	4.6320	0.0048
$\Phi_{\sigma,4}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.8091	2.18	0.0238
Standard Deviation	0.0999	1.2603	0.0046
$\Phi_{\#}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.8180	3.54	0.0234
Standard Deviation	0.0511	0.69	0.0029

Table 6.2

A comparison of rank-1 approximation methods with tolerance parameter $\varepsilon = 10^{-6}$

Random (10 runs)	Max Fit	Total # Iterations	Total Time
Average	0.6665	92.2550	0.1195
Standard Deviation	0.2411	11.4910	0.0138
Quick Rank 1	Fit	# Iterations	Time
Average	0.6788	2.1760	0.0925
Standard Deviation	0.1700	0.8425	0.0114
$\Phi_{\sigma,4}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.7040	2.0790	0.0989
Standard Deviation	0.1579	0.5244	0.0115
$\Phi_{\#}$ and Quick Rank 1	Fit	# Iterations	Time
Average	0.7607	2.0450	0.0989
Standard Deviation	0.1079	0.3809	0.0117

Table 6.3

A comparison of rank 2 approximation methods with tolerance parameter $\varepsilon = 10^{-4}$

7. Conclusion. Colored Brauer diagrams are a graphical way to represent invariant features in tensor data and can be used to visualize calculations with higher order tensors, and analyse the computational complexity of related algorithms. We have used such graphical calculations to find approximations of the spectral norm and to define polynomial maps that amplify the low rank structure of tensors. Such amplification maps are useful for finding better low rank approximations of tensors and are worthy of further study. We are interested in studying n -edge-colored large Brauer diagrams when $n > 3$ and generalizing the given methods for the tensors of order greater than 3. The complexity of computing invariant features corresponding to large diagrams can be high, depending on the particular diagram. In future research, we will investigate how one can improve such computations by using low rank tensor approximations for intermediate results within the calculations.

8. Acknowledgements. This work was partially supported by the National Science Foundation under Grant No. 1837985 and by the Department of Defense under Grant No. BA150235. Neriman Tokcan was partially supported by University of Michigan Precision Health Scholars Grant No. U063159.

REFERENCES

- [1] B. W. Bader, T. G. Kolda and others, *MATLAB Tensor Toolbox, Version 2.6*, available online at <https://www.tensor toolbox.org>, 20XX.
- [2] R. Brauer, *On algebras which are connected with the semisimple continuous groups*, *Annals of Mathematics* **38** (1937), no. 4, 857–872.
- [3] D. Callan, *A combinatorial survey of identities for the double factorial* (2009), arXiv:0906.1317.
- [4] E. J. Candès, B. Recht, *Exact Matrix Completion via Convex Optimization*, *Foundations of Computational Mathematics* **9** (2009), no. 6, <https://doi.org/10.1007/s10208-009-9045-5>.
- [5] E. J. Candès and T. Tao, *The Power of Convex Relaxation: Near-optimal Matrix Completion*, *IEEE Trans. Inf. Theor.* **56** (2010), no. 5, <https://doi.org/10.1109/TIT.2010.2044061>.
- [6] H. Derksen, *On the Nuclear Norm and the Singular Value Decomposition of Tensor*, *Foundations of Computational Mathematics* **16** (2016), no. 3, 779–811.
- [7] S. Friedland, L.-H. Lim, *Nuclear norm of higher-order tensors*, *Mathematics of Computation* **87** (2018), no. 311, 1255–1281.
- [8] R. Goodman and N. R. Wallach, *Representations and Invariants of Classical Groups*, Cambridge University Press, (1998).
- [9] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, *Mem. Amer. Math. Soc.* (1955), no. 16.
- [10] J. Håstad, *Tensor rank is NP-complete*, *Journal of Algorithms* **11** (1990), no. 4, 644–654.
- [11] J. Håstad, *Tensor rank is NP-complete*, *Automata, languages and programming* (Stresa, 1989), *Lecture Notes in Comput. Sci.* **372** (1989), Springer, Berlin, 451–460.
- [12] C. J. Hillar and L. -H. Lim, *Most tensor problems are NP-hard*, *Journal of the ACM* **60** (2013), no. 6, Art. 45.
- [13] F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, *J. Math. Phys.* **6** (1927), no. 1, 164–189.
- [14] F. L. Hitchcock, *Multiple invariants and generalized rank of a p-way matrix or tensor*, *J. Math. Phys.* **7** (1928), no. 1, 39–79.
- [15] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, *SIAM review* **51** (2009), no. 3, 455–500.
- [16] X. Kong, *A concise proof to the spectral and nuclear norm bounds through tensor partitions*, *Open Mathematics* **17** (2019), 365–373.
- [17] J. M. Landsberg, *Tensors: Geometry and Applications*, *American Mathematical Society* **128** (2012).
- [18] S. Lang, *Algebra*, *Graduate Texts in Mathematics* **211** (2002), 3rd ed., Springer-Verlag, New York.
- [19] L. De Lathauwer, B. De Moor, *A Multilinear Singular Value Decomposition*, *SIAM Journal on Matrix Analysis and Applications* **21** (2000), no.4, 1253–1278, <https://doi.org/10.1137/S0895479896305696>.
- [20] Z. Li, *The spectral norm and the nuclear norm of a tensor based on tensor partitions*, *Matrix Anal. Appl.* **37** (2016), no. 4, 1440–1452.
- [21] Z. Li, Y. Nakatsukasa, and others, *On Orthogonal Tensors and Best Rank-One Approximation Ratio*, *SIAM Journal on Matrix Analysis and Applications* **39** (2017), <https://doi.org/10.1137/17M1144349>.
- [22] OEIS Foundation Inc. (2020), *The On-Line Encyclopedia of Integer Sequences*, <http://oeis.org/A001147>.
- [23] OEIS Foundation Inc. (2019), *The On-Line Encyclopedia of Integer Sequences*, <http://oeis.org/A002831>.
- [24] V. L. Popov and E. B. Vinberg, “Invariant theory,” in: *Encyclopaedia of Mathematical Sciences* **55** (1994), A. Parshin and I. R. Shafarevich, eds., Springer-Verlag, Berlin.
- [25] L. Qi and S. Hu, *Spectral Norm and Nuclear Norm of a Third Order Tensor* (2019), arXiv:1909.01529.
- [26] A. Ramlatchan, M. Yang, and others, *A survey of matrix completion methods for recommendation systems*, *Big Data Mining and Analytics* **1** (2018), no. 4, 308–323.

- [27] R. Schatten, *A Theory of Cross-Spaces*, Princeton University Press (1950), Princeton, NJ.
- [28] A. P. Da Silva, P. Comon, and others, *A Finite Algorithm to Compute Rank-1 Tensor Approximations*, IEEE Signal Processing Letters **23** (2016), no. 7, 959–963.
- [29] V. De Silva and L. -H.Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM Journal on Matrix Analysis and Applications **30** (2008), 10841127.
- [30] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika **31** (1966), 279-311, <https://doi.org/10.1007/BF02289464>.
- [31] H. Weyl, *The Classical Groups. Their Invariants and Representations* (1939), Princeton University Press.
- [32] L. K. Williams, *Invariant Polynomials on Tensors under the Action of a Product of Orthogonal Groups*, Ph.D. Thesis, University of Wisconsin–Milwaukee, 2013.
- [33] M. Yuan and C. Zhang, *On Tensor Completion via Nuclear Norm Minimization*, Foundations of Computational Mathematics **16** (2016), no. 4, 1031–1068, <https://doi.org/10.1007/s10208-015-9269-5>.
- [34] T. Zhang and G. H. Golub, *Rank-One Approximation to High Order Tensors*, SIAM Journal on Matrix Analysis and Applications **23** (2001), no. 2, 534–550, <https://doi.org/10.1137/S0895479899352045>.