How Jellyfish Characterise Alternating Group Equivariant Neural Networks

Edward Pearce–Crump [1]

Abstract

We provide a full characterisation of all of the possible alternating group (A_n) equivariant neural networks whose layers are some tensor power of $\mathbb{R}^n$. In particular, we find a basis of matrices for the learnable, linear, A_n–equivariant layer functions between such tensor power spaces in the standard basis of $\mathbb{R}^n$. We also describe how our approach generalises to the construction of neural networks that are equivariant to local symmetries.

1. Introduction

There has been a growing research effort in deep learning to develop neural network architectures that can be used to learn efficiently from data that possesses an underlying symmetry. These architectures guarantee that the functions that are learned are subject to a geometric property, known as *equivariance*, that is connected with the symmetry group. Group equivariant neural networks are important due to the additional advantages that they offer over traditional multilayer perceptron models. For example, they commonly display high levels of parameter sharing within each layer, resulting in significantly fewer parameters overall. This often leads to models that show improved prediction performance on new data.

The symmetry group that has received the most attention, in terms of it being explicitly incorporated into neural network architectures, is the group of all permutations on some fixed number of objects, called the symmetric group. Creating neural networks that are equivariant to permutations is highly desirable as many data structures, such as sets and graphs, exhibit natural permutation symmetry. It is easily understood that the labelling of the elements of a set or the vertices of a graph is arbitrary; hence, it is crucial to ensure that the functions that are learned from such data do not depend on how the data is labelled.

[1]Department of Computing, Imperial College London, United Kingdom. Correspondence to: Edward Pearce–Crump <ep1011@ic.ac.uk>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

In this paper, we look instead at how to construct neural networks that are equivariant to the alternating group. The alternating group is an index two subgroup of the symmetric group consisting solely of all of the even permutations. Alternating group symmetry has proven to be particularly useful when learning from spherical image data that has been discretely represented on an icosahedron (Zhang et al., 2019); in constructing convolutional neural networks on an icosahedron (Cohen et al., 2019); and in estimating polynomials that are invariant to the action of the alternating group (Kicki et al., 2020).

Specifically, we give a full characterisation of all of the possible alternating group equivariant neural networks whose layers are some tensor power of $\mathbb{R}^n$ by finding a basis of matrices for the learnable, linear, alternating group equivariant layer functions between such tensor power spaces in the standard basis of $\mathbb{R}^n$.

Our approach is similar the one presented in the papers written by Pearce–Crump (2022a; 2022b). They used different sets of set partition diagrams to characterise all of the learnable, linear, group equivariant layer functions between tensor power spaces in the standard basis of $\mathbb{R}^n$ for the following groups: the symmetric group S_n; the orthogonal group $O(n)$; the symplectic group $Sp(n)$; and the special orthogonal group $SO(n)$. We will show that in the case of the alternating group A_n, the layer functions can also be characterised by certain sets of set partition diagrams.

To do this, we use a concept that was first introduced by Comes (2020), namely, so-called *jellyfish*. In their paper, they largely determined the theory of alternating group equivariance; however, they relied heavily on the language of category theory in their exposition. We simplify their approach, and provide proofs that are more accessible to the machine learning community.

The main contributions of this paper, which appear in Section 6 onwards, are as follows:

1. We are the first to show how the combinatorics underlying set partition diagrams, together with some jellyfish, serves as the theoretical foundation for constructing neural networks that are equivariant to the alternating group when the layers are some tensor power of $\mathbb{R}^n$.

2. In particular, we find a basis for the learnable, linear, A_n–equivariant layer functions between such tensor power spaces in the standard basis of $\mathbb{R}^n$.

3. We extend our approach to show how to construct neural networks that are equivariant to local symmetries.

2. Preliminaries

We choose our field of scalars to be $\mathbb{R}$ throughout. Tensor products are also taken over $\mathbb{R}$, unless otherwise stated. Also, we let $[n]$ represent the set $\{1, \dots, n\}$.

Recall that a representation of a group G is a choice of vector space V over $\mathbb{R}$ and a group homomorphism

$$\rho : G \to GL(V) \tag{1}$$

We choose to focus on finite-dimensional vector spaces V that are some tensor power of $\mathbb{R}^n$ in this paper.

We often abuse our terminology by calling V a representation of G, even though the representation is technically the homomorphism ρ. When the homomorphism ρ needs to be emphasised alongside its vector space V, we will use the notation (V, ρ).

3. $(\mathbb{R}^n)^{\otimes k}$, a representation of both S_n and A_n

Recall that S_n is the group of all permutations on $[n]$, and that A_n is the subgroup of S_n consisting of all permutations on $[n]$ whose image under the function

$$\mathrm{sgn} : S_n \to \{\pm 1\} \tag{2}$$

is $+1$, where sgn is defined, for all $\sigma \in S_n$, as

$$\mathrm{sgn}(\sigma) \coloneqq \begin{cases} +1 & \text{if } \sigma \text{ is an even permutation in } S_n \\ -1 & \text{if } \sigma \text{ is an odd permutation in } S_n \end{cases} \tag{3}$$

We have that $\mathbb{R}^n$ is a representation of S_n via its (left) action on the basis $\{e_a \mid a \in [n]\}$ which is extended linearly, where, specifically, the action is given by

$$\sigma \cdot e_a = e_{\sigma(a)} \text{ for all } \sigma \in S_n \text{ and } a \in [n] \tag{4}$$

Restricting this action to A_n and extending linearly shows that $\mathbb{R}^n$ is also a representation of A_n.

We also have that any k-tensor power of $\mathbb{R}^n$, $(\mathbb{R}^n)^{\otimes k}$, for any $k \in \mathbb{Z}_{\geq 0}$, is a representation of S_n, since the elements

$$e_I \coloneqq e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k} \tag{5}$$

for all $I \coloneqq (i_1, i_2, \dots, i_k) \in [n]^k$ form a basis of $(\mathbb{R}^n)^{\otimes k}$, and the action of S_n that maps a basis element of $(\mathbb{R}^n)^{\otimes k}$ of the form (5) to

$$e_{\sigma(I)} \coloneqq e_{\sigma(i_1)} \otimes e_{\sigma(i_2)} \otimes \cdots \otimes e_{\sigma(i_k)} \tag{6}$$

can be extended linearly. Again, restricting this action to A_n and extending linearly shows that $(\mathbb{R}^n)^{\otimes k}$ is also a representation of A_n.

We denote the representation of S_n by ρ_k. We will use the same notation for the restriction of this representation to A_n, with the context making clear that it is the restriction of the S_n representation.

For more on the representation theory of the symmetric and alternating groups, see (Sagan, 2000) and (Ceccherini-Silberstein et al., 2010).

4. Group Equivariant Neural Networks

Group equivariant neural networks are constructed by alternately composing linear and non-linear G-equivariant maps between representations of a group G. The following is based on the material presented in (Lim & Nelson, 2022).

We first define *G-equivariance*:

Definition 4.1. Suppose that (V, ρ_V) and (W, ρ_W) are two representations of a group G.

A map $\phi : V \to W$ is said to be G-equivariant if, for all $g \in G$ and $v \in V$,

$$\phi(\rho_V(g)[v]) = \rho_W(g)[\phi(v)] \tag{7}$$

The set of all *linear* G-equivariant maps between V and W is denoted by $\mathrm{Hom}_G(V, W)$. When $V = W$, we write this set as $\mathrm{End}_G(V)$. It can be shown that $\mathrm{Hom}_G(V, W)$ is a vector space over $\mathbb{R}$, and that $\mathrm{End}_G(V)$ is an algebra over $\mathbb{R}$. See (Segal, 2014) for more details.

A special case of G-equivariance is *G-invariance*:

Definition 4.2. The map ϕ given in Definition 4.1 is said to be G-invariant if ρ_W is defined to be the 1-dimensional trivial representation of G. As a result, $W = \mathbb{R}$.

We can now define the type of neural network that is the focus of this paper:

Definition 4.3. An L-layer G-equivariant neural network f_{NN} is a composition of *layer functions*

$$f_{NN} \coloneqq f_L \circ \ldots \circ f_l \circ \ldots \circ f_1 \tag{8}$$

such that the lth layer function is a map of representations of G

$$f_l : (V_{l-1}, \rho_{l-1}) \to (V_l, \rho_l) \tag{9}$$

that is itself a composition

$$f_l \coloneqq \sigma_l \circ \phi_l \tag{10}$$

of a learnable, linear, G-equivariant function

$$\phi_l : (V_{l-1}, \rho_{l-1}) \to (V_l, \rho_l) \tag{11}$$

together with a fixed, non-linear activation function

$$\sigma_l : (V_l, \rho_l) \to (V_l, \rho_l) \tag{12}$$

such that

1. σ_l is a G-equivariant map, as in (7), and
2. σ_l acts pointwise (after a basis has been chosen for each copy of V_l in σ_l.)

We focus on the learnable, linear, G-equivariant functions in this paper, since the non-linear functions are fixed. Note that, after picking a basis for each layer space in the set $\{V_l\}$, the number of parameters that appear in a matrix representation of a learnable, linear, G-equivariant function of the form (11), that is, in a weight matrix for the lth layer function, is equal to the number of matrices that appear in a basis for $\mathrm{Hom}_G(V_{l-1}, V_l)$. Furthermore, given a basis of $\mathrm{Hom}_G(V_{l-1}, V_l)$, the weight matrix itself is a weighted linear combination of these basis matrices, where each coefficient in the linear combination is a parameter to be learned.

Remark 4.4. The entire neural network f_{NN} is itself a G-equivariant function because it can be shown that the composition of any number of G-equivariant functions is itself G-equivariant.

Remark 4.5. One way of making a neural network of the form given in Definition 4.3 G-invariant is by choosing the representation in the final layer to be the 1-dimensional trivial representation of G.

5. Symmetric Group

In this section, we recall the technique of using set partitions to find a basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ in the standard basis of $\mathbb{R}^n$ since it will feature heavily in what follows for the alternating group. For more details, see (Maron et al., 2019a), (Ravanbakhsh, 2020), and (Pearce-Crump, 2022a).

5.1. Set Partitions

For $l, k \in \mathbb{Z}_{\geq 0}$, consider the set $[l+k]$ having $l+k$ elements. We can create a set partition of $[l+k]$ by partitioning it into a number of subsets. We call the subsets of a set partition *blocks*. Define Π_{l+k} to be the set of all set partitions of $[l+k]$. Let $\Pi_{l+k,n}$ be the subset of Π_{l+k} consisting of all set partitions of $[l+k]$ having at most n blocks.

As the number of set partitions in Π_{l+k} having exactly t blocks is the Stirling number $\left\{ {l+k \atop t} \right\}$ of the second kind, we see that the number of elements in Π_{l+k} is equal to $\mathrm{B}(l+k)$, the $(l+k)^{\text{th}}$ Bell number, and that the number of elements in $\Pi_{l+k,n}$ is therefore equal to

$$\sum_{t=1}^{n} \left\{ {l+k \atop t} \right\} := \mathrm{B}(l+k, n) \tag{13}$$

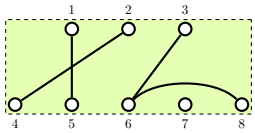


Figure 1. A diagram d_π corresponding to the set partition π given in (14) with $l = 3$ and $k = 5$.

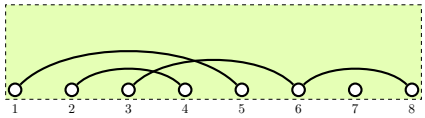


Figure 2. The flattened diagram corresponding to the set partition diagram shown in Figure 1.

the n-restricted $(l+k)^{\text{th}}$ Bell number.

For each set partition π in Π_{l+k}, we can define a diagram d_π that has two rows of vertices and edges between vertices such that there are

1. l vertices on the top row, labelled by $1, \dots, l$
2. k vertices on the bottom row, labelled by $l+1, \dots, l+k$, and
3. the edges between the vertices are such that the connected components of d_π correspond to the blocks of π.

In particular, d_π represents the equivalence class of all diagrams with connected components equal to the blocks of π.

For example, if $l = 3$ and $k = 5$, a diagram corresponding to the set partition

$$\pi := \{1, 5 \mid 2, 4 \mid 3, 6, 8 \mid 7\} \tag{14}$$

in Π_8 consisting of 4 blocks is given in Figure 1.

We can define another diagram related to each set partition diagram d_π: it is the flattened version of d_π, where we pull the top row of l vertices down and to the left of the bottom row of k vertices, maintaining the order of the labels.

For example, the flattened diagram corresponding to the set partition diagram shown in Figure 1 is given in Figure 2.

5.2. A Basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$

We begin by noting that $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ has a standard basis of matrix units

$$\{E_{I,J}\}_{I \in [n]^l, J \in [n]^k} \tag{15}$$

where $E_{I,J}$ has a 1 in the (I, J) position and is 0 elsewhere.

Hence, expressing $f \in \mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ in this basis as

$$f = \sum_{I \in [n]^l} \sum_{J \in [n]^k} f_{I,J} E_{I,J} \tag{16}$$

it can be shown that $f \in \mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ if and only if

$$f_{\sigma(I),\sigma(J)} = f_{I,J} \tag{17}$$

for all $\sigma \in S_n$ and $I \in [n]^l, J \in [n]^k$.

Concatenating each $I \in [n]^l, J \in [n]^k$ into a single element $(I, J) \in [n]^{l+k}$, we see from (17) that the basis elements of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ are in bijective correspondence with the orbits coming from the action of S_n on $[n]^{l+k}$, where $\sigma \in S_n$ acts on the pair (I, J) by

$$\sigma(I, J) \coloneqq (\sigma(I), \sigma(J)) \tag{18}$$

As S_n acts on $[n]$ transitively, it acts on $[n]^{l+k}$ transitively, meaning that the orbits under this action actually partition the set $[n]^{l+k}$ into equivalence classes.

Furthermore, we can define a bijection between the orbits coming from the action of S_n on $[n]^{l+k}$ and the set partitions π in Π_{l+k} having at most n blocks. Indeed, if (I, J) is a class representative of an orbit, then, replacing momentarily the elements of J by $i_{l+m} \coloneqq j_m$ for all $m \in [k]$, so that

$$\begin{aligned}(I, J) &= (i_1, i_2, \dots, i_l, j_1, j_2, \dots, j_k) \\ &= (i_1, i_2, \dots, i_l, i_{l+1}, i_{l+2}, \dots i_{l+k})\end{aligned} \tag{19}$$

we define the bijection by

$$i_x = i_y \iff x, y \text{ are in the same block of } \pi \tag{20}$$

for all $x, y \in [l + k]$.

The bijection (20) is independent of the choice of class representative since

$$i_x = i_y \iff \sigma(i_x) = \sigma(i_y) \text{ for all } \sigma \in S_n \tag{21}$$

Notice that the LHS of (20) is checking for an equality on the elements of $[n]$, whereas the RHS is separating the elements of $[l + k]$ into blocks; hence π must have at most n blocks.

As a result, we have shown the following.

Theorem 5.1. *The basis elements of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, in the standard basis of $\mathbb{R}^n$, correspond bijectively with all set partitions π in Π_{l+k} having at most n blocks, which correspond bijectively with the orbits coming from the action of S_n on $[n]^{l+k}$.*

We can obtain the basis elements themselves, as follows:

Taking a set partition π in $\Pi_{l+k,n}$, and denoting the number of blocks in π by t, we obtain a labelling of the blocks by letting B_1 be the block that contains the number $1 \in [l + k]$, and iteratively letting B_j, for $1 < j \leq t$, be the block that contains the smallest number in $[l + k]$ that is not in $B_1 \cup B_2 \cup \dots \cup B_{j-1}$.

Using this labelling of the blocks, we can create an element of $[n]^{l+k}$ by letting the ith position, $i \in [l + k]$, be the label of the block containing the number i. This is called the *block labelling* of π. Denote it by the pair (I_π, J_π), where clearly $I_\pi \in [n]^l$ and $J_\pi \in [n]^k$.

In doing so, we have created a representative of the orbit for the S_n action on $[n]^{l+k}$ corresponding to the set partition $\pi \in \Pi_{l+k,n}$ under the bijection given in (20). Denote this orbit by $O_{S_n}((I_\pi, J_\pi))$.

We form a basis element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, denoted by X_π, by adding together all matrix units whose indexing pair (I, J) appears in the orbit $O_{S_n}((I_\pi, J_\pi))$; that is,

$$X_\pi \coloneqq \sum_{(I,J) \in O_{S_n}((I_\pi, J_\pi))} E_{I,J} \tag{22}$$

We see that X_π is a basis element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ by (17).

Doing this for each set partition π in $\Pi_{l+k,n}$, or, equivalently, for all of the orbits coming from the action of S_n on $[n]^{l+k}$, gives:

Theorem 5.2. *For $l, k \in \mathbb{Z}_{\geq 0}, n \in \mathbb{Z}_{\geq 1}$, we have that*

$$\{X_\pi \mid \pi \in \Pi_{l+k,n}\} \tag{23}$$

is a basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, and so

$$\dim \mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l}) = \mathrm{B}(l + k, n) \tag{24}$$

6. Alternating Group

In exactly the same way as for the symmetric group, we can show that the basis elements of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ are in bijective correspondence with the orbits coming from the action of A_n on $[n]^{l+k}$.

However, the major difference between the symmetric group and the alternating group is that the A_n orbits on $[n]^{l+k}$, and consequently the basis elements of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, are not necessarily in bijective correspondence with the set partitions of $[l + k]$ having at most n blocks. This is because some of the S_n orbits on $[n]^{l+k}$ become the disjoint union of more than one A_n orbit on $[n]^{l+k}$. We say that an S_n orbit on $[n]^{l+k}$ *splits* if it is the disjoint union of more than one A_n orbit on $[n]^{l+k}$. Our first task is to identify which S_n orbits on $[n]^{l+k}$ split and which do not.

Theorem 6.1 (S_n orbits split)**.** *Let π be a set partition in $\Pi_{l+k,n}$ having some t blocks, and consider its corresponding S_n orbit on $[n]^{l+k}$, $O_{S_n}((I_\pi, J_\pi))$.*

If $t \leq n-2$, then $O_{S_n}((I_\pi, J_\pi))$ does not split.

If $t = n-1$ or n, then $O_{S_n}((I_\pi, J_\pi))$ splits into a disjoint union of exactly two A_n orbits.

Proof. Consider the case where π has $t = n$ blocks. Choose some arbitrary element $(I_\alpha, J_\alpha) \in O_{S_n}((I_\pi, J_\pi))$, and consider $\mathrm{Stab}_{A_n}((I_\alpha, J_\alpha))$, the stabilizer of (I_α, J_α) under the action of A_n. Since all of the elements in $[n]$ appear in the tuple (I_α, J_α), the only element of A_n that fixes the entries of (I_α, J_α) is the identity element; that is, $\mathrm{Stab}_{A_n}((I_\alpha, J_\alpha)) = \{e_{A_n}\}$. Hence, by the Orbit–Stabilizer Theorem, we have that $|O_{A_n}((I_\alpha, J_\alpha))| = \frac{n!}{2}$. Since $|O_{S_n}((I_\pi, J_\pi))| = n!$ in this case, and because (I_α, J_α) was an arbitary element of $O_{S_n}((I_\pi, J_\pi))$, we see that, by another application of the Orbit–Stabilizer Theorem, $O_{S_n}((I_\pi, J_\pi))$ splits into a disjoint union of two A_n orbits.

Similarly, for the case where π has $t = n-1$ blocks, we see that $\mathrm{Stab}_{A_n}((I_\alpha, J_\alpha)) = \{e_{A_n}\}$, and so, by the same argument, $O_{S_n}((I_\pi, J_\pi))$ splits into a disjoint union of two A_n orbits.

Finally, for $t \leq n-2$, we have that $\mathrm{Stab}_{A_n}((I_\alpha, J_\alpha)) \cong A_{n-t}$, and so, by the Orbit–Stabilizer Theorem, we see that $|O_{A_n}((I_\alpha, J_\alpha))| = |O_{S_n}((I_\pi, J_\pi))|$. Consequently, the S_n orbit $O_{S_n}((I_\pi, J_\pi))$ does not split. □

Remark 6.2. The case for $t \leq n-2$ in Theorem 6.1 was proven independently in (Maron et al., 2019b).

As a result, we immediately obtain the following two theorems:

Theorem 6.3. *Let $k, l \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{Z}_{\geq 1}$.*

If π is a set partition in $\Pi_{l+k,n}$ having $n-2$ blocks or fewer, then π corresponds bijectively to a basis element of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$,

Otherwise, π corresponds to two basis elements of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$.

Theorem 6.4. *For $k, l \in \mathbb{Z}_{\geq 0}, n \in \mathbb{Z}_{\geq 1}$, the dimension of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ is equal to*

$$\sum_{t=1}^{n-2} \left\{ {l+k \atop t} \right\} + 2 \left\{ {l+k \atop n-1} \right\} + 2 \left\{ {l+k \atop n} \right\} \tag{25}$$

In other words, Theorem 6.3 tells us that finding a basis of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ in the standard basis of $\mathbb{R}^n$ is very similar to finding a basis of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ in the standard basis of $\mathbb{R}^n$, since finding the basis amounts once again to considering all of the set partitions of $[l+k]$ having at most n blocks.

Indeed, if a set partition π in $\Pi_{l+k,n}$ has at most $n-2$ blocks, then we see that X_π given in (22) is a basis element of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ since, in this case, $O_{S_n}((I_\pi, J_\pi)) = O_{A_n}((I_\pi, J_\pi))$, by Theorem 6.1.

The question remains as to how to take a set partition π in $\Pi_{l+k,n}$ having either $n-1$ or n blocks and use it to obtain the two basis elements of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ that it corresponds to. Said differently, we would like to take such a set partition π and identify the two A_n orbits that its corresponding S_n orbit on $[n]^{l+k}$, $O_{S_n}((I_\pi, J_\pi))$, splits into.

6.1. Enter: the Jellyfish

Comes (2020) originally came up with the idea of using *jellyfish* to identify the two A_n orbits; here, however, our choice of exposition is rather different and our proofs are simpler than those given in their paper.

We begin by defining the following map.

Definition 6.5. For $n \in \mathbb{Z}_{\geq 1}$, we define the determinant map

$$\det : (\mathbb{R}^n)^{\otimes n} \to \mathbb{R} \tag{26}$$

on the standard basis of $(\mathbb{R}^n)^{\otimes n}$ by

$$e_I = e_{i_1} \otimes \cdots \otimes e_{i_n} \mapsto \begin{vmatrix} e_{i_1} & \dots & e_{i_n} \end{vmatrix} \tag{27}$$

and extend linearly, where the RHS of (27) is the determinant of an $n \times n$ matrix.

Lemma 6.6. *The determinant map is an element of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes n}, (\mathbb{R}^n)^{\otimes 0})$, but it is not an element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes n}, (\mathbb{R}^n)^{\otimes 0})$.*

Proof. It is clear that the determinant map is a linear map.

Then, for any $\sigma \in S_n$, we have that

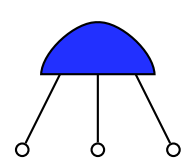
$$\det(\rho_n(\sigma)[e_I]) = (-1)^{\mathrm{sgn}(\sigma)} \det(e_I) \tag{28}$$

since a transposition in S_n corresponds to swapping two columns of the $n \times n$ matrix.

As $\mathrm{sgn}(\sigma) = 1$ for all $\sigma \in A_n$, (28) shows that $\det$ is an element of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes n}, (\mathbb{R}^n)^{\otimes 0})$. It is enough to see that for any odd permutation in S_n, (28) implies that $\det$ is not an element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes n}, (\mathbb{R}^n)^{\otimes 0})$. □

For each $n \in \mathbb{Z}_{\geq 1}$, we represent the determinant map by a diagram that has a single row of n vertices, each of which is attached to a blue head. We will call this diagram a *jellyfish*, for obvious reasons.

For example, if $n = 3$, the jellyfish has the form



Remark 6.7. It is important to highlight that the determinant map (26) is different from the determinant operator on an $n \times n$ matrix. In particular, the determinant map is a linear map on $(\mathbb{R}^n)^{\otimes n}$ whereas the determinant operator on $n \times n$ matrices is not linear.

We introduce the following useful lemma.

Lemma 6.8. *The determinant map applied to a standard basis vector of* $(\mathbb{R}^n)^{\otimes n}$ *gives either* $-1, 0$ *or* $+1$.

Proof. Let $e_I = e_{i_1} \otimes \cdots \otimes e_{i_n}$ be a standard basis vector of $(\mathbb{R}^n)^{\otimes n}$.

If $i_x = i_y$ for some $1 \leq x \neq y \leq n$, then $\det(e_I) = 0$ since two columns in the $n \times n$ matrix will be the same. Otherwise, $\{i_1, \dots, i_n\}$ is some permutation of $[n]$. Defining $e_{[n]} := e_1 \otimes \cdots \otimes e_n$, we know that $\det(e_{[n]}) = +1$. Since, in this case, $e_I = \rho_n(\sigma) e_{[n]}$ for some $\sigma \in S_n$, we can use (28) to calculate $\det(e_I)$, giving a result of ± 1. □

Remark 6.9. We can view Lemma 6.8 in another way; namely, that the determinant map splits basis vectors of $(\mathbb{R}^n)^{\otimes n}$ into disjoint classes. Let us call these classes $-1, 0$ and $+1$.

Consequently, Remark 6.9 suggests that the determinant map is a possible candidate function for identifying the A_n orbits that the S_n orbit $O_{S_n}((I_\pi, J_\pi))$ splits into, where π is a set partition in $\Pi_{l+k,n}$ having either $n-1$ or n blocks.

However, since the determinant map is a map $(\mathbb{R}^n)^{\otimes n} \to \mathbb{R}$, and the elements of $O_{S_n}((I_\pi, J_\pi))$ are elements of $[n]^{l+k}$, to use the determinant map to try to identify the A_n orbits in $O_{S_n}((I_\pi, J_\pi))$, it would be useful to create a map $g_\pi : (\mathbb{R}^n)^{\otimes l+k} \to (\mathbb{R}^n)^{\otimes n}$ that corresponds bijectively with the set partition π, since such a map would project standard basis elements of $(\mathbb{R}^n)^{\otimes l+k}$ onto, at the very least, a linear combination of basis elements of $(\mathbb{R}^n)^{\otimes n}$. We could then use the splitting property of the determinant map on such linear combinations to try to split the elements of $O_{S_n}((I_\pi, J_\pi))$ into different classes. However, in order to use the three classes given in Remark 6.9, we cannot choose any map $(\mathbb{R}^n)^{\otimes l+k} \to (\mathbb{R}^n)^{\otimes n}$. We will see that with a clever choice of g_π, the determinant map then splits the elements of $O_{S_n}((I_\pi, J_\pi))$ into the ± 1 classes and sends all other elements of $[n]^{l+k}$ to the 0 class.

We create g_π as follows. Flatten the diagram d_π that corresponds to π. Add a new top row of n vertices above the row of $l+k$ vertices. From the bottom row, take the lowest numbered vertex in block i of π, $1 \leq i \leq t$, where $t = n-1$ or n, and connect that vertex to vertex i in the top row. Call this new diagram b_π.

We claim the following:

Proposition 6.10. *The diagram* b_π *corresponds bijectively with a basis element of* $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes l+k}, (\mathbb{R}^n)^{\otimes n})$. *Consequently, we define* g_π *to be this basis element.*

Proof. It is clear that the diagram b_π corresponds to a set partition of $[n+l+k]$ having exactly n blocks. Hence this set partition is an element of $\Pi_{n+l+k,n}$. Since each set partition in $\Pi_{n+l+k,n}$ corresponds bijectively with a basis element of $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes l+k}, (\mathbb{R}^n)^{\otimes n})$, by Theorem 5.2, we obtain the result. □

The following result is immediate from the construction of the diagram b_π.

Proposition 6.11. *In the standard basis of matrix units of* $\mathrm{Hom}((\mathbb{R}^n)^{\otimes l+k}, (\mathbb{R}^n)^{\otimes n})$, g_π *is given by*

$$g_\pi := \sum_{(K,I,J) \in O_{S_n}((K_\pi, I_\pi, J_\pi))} E_{K,(I,J)} \tag{29}$$

where

$$K_\pi := (1, 2, \dots, n) \tag{30}$$

and $O_{S_n}((K_\pi, I_\pi, J_\pi))$ *is defined to be*

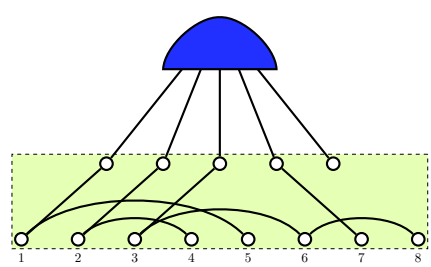
$$\left\{ (K, I, J) \;\middle|\; \begin{array}{l} (I, J) \in O_{S_n}((I_\pi, J_\pi)) \\ \textit{and if } (I, J) = \sigma(I_\pi, J_\pi), \\ \textit{then } K := \sigma(K_\pi) \end{array} \right\} \tag{31}$$

In order to see that g_π has the properties that we need, we first define the following map.

Definition 6.12. Let π be a set partition in $\Pi_{l+k,n}$ having either $n-1$ or n blocks. Then we define $f_\pi := \det \circ g_\pi : (\mathbb{R}^n)^{\otimes l+k} \to \mathbb{R}$. Clearly, f_π corresponds bijectively with the set partition π.

We can associate to f_π a diagram that is built from the composition of b_π (which corresponds to the map g_π) and a jellyfish (which corresponds to the determinant map).

For example, the diagram that is associated to f_π for the set partition π whose diagram d_π is given in Figure 1, where we choose $l = 3$, $k = 5$ and $n = 5$, is



Proposition 6.13. *The map* f_π *is an element of* $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes l+k}, (\mathbb{R}^n)^{\otimes 0})$.

Proof. Let $e_{(I,J)}$ be any standard basis vector in $(\mathbb{R}^n)^{\otimes l+k}$, where $I \in [n]^l$ and $J \in [n]^k$.

Then, for all $\sigma \in A_n$, we have that

$$f_\pi(\rho_{l+k}(\sigma)[e_{(I,J)}]) = \det \circ g_\pi(\rho_{l+k}(\sigma)[e_{(I,J)}]) \quad (32)$$
$$= \det(\rho_n(\sigma)[g_\pi(e_{(I,J)})]) \quad (33)$$
$$= \det(g_\pi(e_{(I,J)})) \quad (34)$$
$$= f_\pi(e_{(I,J)}) \quad (35)$$

where, in (33), we have used Proposition 6.10 and in (34), we have used (28). □

We come to the crucial result for our purposes.

Theorem 6.14. *Let $e_{(I,J)}$ be any standard basis vector in $(\mathbb{R}^n)^{\otimes l+k}$, where $I \in [n]^l, J \in [n]^k$.*

Then

$$f_\pi : e_{(I,J)} \to \begin{cases} \pm 1 & \text{if } (I,J) \in O_{S_n}((I_\pi, J_\pi)) \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

Proof. For any standard basis vector $e_{(I,J)}$ in $(\mathbb{R}^n)^{\otimes l+k}$, we have, on a matrix unit of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes l+k}, (\mathbb{R}^n)^{\otimes n})$, that

$$E_{K,(L,M)} e_{(I,J)} = \delta_{((L,M),(I,J))} e_K \quad (37)$$

where e_K is a standard basis vector in $(\mathbb{R}^n)^{\otimes n}$.

Hence, by (29), we see that, for the otherwise case, $g_\pi(e_{(I,J)}) = 0$, and so $f_\pi(e_{(I,J)}) = 0$.

If $(I,J) \in O_{S_n}((I_\pi, J_\pi))$, then we have that $(I,J) = \sigma(I_\pi, J_\pi)$ for some $\sigma \in S_n$.

Hence, by (29), (37) and the definition of $O_{S_n}((K_\pi, I_\pi, J_\pi))$ given in (31), we see that $g_\pi(e_{(I,J)}) = e_K$, where $K = \sigma(K_\pi)$.

Since K is a permutation of $[n]$, we can apply Lemma 6.8 to get that $f_\pi(e_{(I,J)}) = \det(e_K) = \pm 1$. □

Consequently, we can define the following two sets:

$$O_\pi^+ := \{(I,J) \in O_{S_n}((I_\pi, J_\pi)) \mid f_\pi(e_{(I,J)}) = +1\} \quad (38)$$

and

$$O_\pi^- := \{(I,J) \in O_{S_n}((I_\pi, J_\pi)) \mid f_\pi(e_{(I,J)}) = -1\} \quad (39)$$

We claim the following result.

Theorem 6.15. *O_π^+ and O_π^- are the two A_n orbits that $O_{S_n}((I_\pi, J_\pi))$ splits into.*

Proof. It is enough to show that if $(I,J) \in O_\pi^+$, then $\sigma(I,J) \in O_\pi^+$ for all $\sigma \in A_n$, and if $(I,J) \in O_\pi^-$, then $\sigma(I,J) \in O_\pi^-$ for all $\sigma \in A_n$, where the action of σ on the pair (I,J) is given in (18). Indeed, by Proposition 6.13, we have that

$$f_\pi(e_{\sigma(I,J)}) = f_\pi(\rho_{l+k}(\sigma)[e_{(I,J)}]) = f_\pi(e_{(I,J)}) \quad (40)$$

Applying Theorem 6.14 gives the result. □

Relabelling O_π^+ as $O_{A_n}((I_\pi^+, J_\pi^+))$, and O_π^- as $O_{A_n}((I_\pi^-, J_\pi^-))$, we obtain the two basis elements of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ from the one set partition $\pi \in \Pi_{l+k,n}$ that has either $n-1$ or n blocks, namely

$$X_\pi^+ := \sum_{(I,J) \in O_{A_n}((I_\pi^+, J_\pi^+))} E_{I,J} \quad (41)$$

and

$$X_\pi^- := \sum_{(I,J) \in O_{A_n}((I_\pi^-, J_\pi^-))} E_{I,J} \quad (42)$$

We summarise the above results into the following theorem.

Theorem 6.16. *For $l, k \in \mathbb{Z}_{\geq 0}, n \in \mathbb{Z}_{\geq 1}$, the union of the two sets*

$$\{X_\pi \mid \pi \in \Pi_{l+k,n-2}\} \quad (43)$$

$$\{X_\pi^+, X_\pi^- \mid \pi \in \Pi_{l+k,n} \setminus \Pi_{l+k,n-2}\} \quad (44)$$

forms a basis of $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$. Its dimension is given in Theorem 6.4.

In Algorithm 1, we present some pseudocode for how to explicitly construct the weight matrix for an A_n-equivariant linear layer mapping $(\mathbb{R}^n)^{\otimes k} \to (\mathbb{R}^n)^{\otimes l}$ in the standard basis of $\mathbb{R}^n$. We assume that we have access to the following procedures:

- SYMMGRPORBIT calculates the S_n orbit for a set partition $\pi \in \Pi_{l+k,n}$.
- FLATTEN takes a set partition diagram with l vertices in the top row and k vertices in the bottom row, and returns its equivalent flattened set partition diagram having a single row of $l+k$ vertices, as per Section 5.1.
- NEWTOPROWCONNECT takes a flattened set partition diagram having either $n-1$ or n blocks, inserts a new top row consisting of n vertices, and connects the lowest numbered vertex in each block i to vertex i in the top row.
- ATTACHJELLYFISH takes a set partition diagram b_π with n vertices in the top row and $l+k$ vertices in the bottom row having either $n-1$ or n blocks, attaches a jellyfish with n legs to the top row, and returns the function f_π that is associated with this new diagram.

Algorithm 1 How to Calculate the Weight Matrix for an A_n Equivariant Linear Layer Mapping $(\mathbb{R}^n)^{\otimes k} \to (\mathbb{R}^n)^{\otimes l}$

Input: n, k, l
Output: Weight Matrix M
Initialize $M = 0$.
for all $\pi \in \Pi_{l+k,n-2}$ **do**
  $O_{S_n}((I_\pi, J_\pi)) = \text{SymmGrpOrbit}(\pi, n)$
  $X_\pi = \sum_{(I,J)\in O_{S_n}((I_\pi,J_\pi))} E_{I,J}$
  $M = M + \lambda_\pi X_\pi$
end for
for all $\pi \in \Pi_{l+k,n} \setminus \Pi_{l+k,n-2}$ **do**
  $d_\pi = \text{Flatten}(d_\pi)$
  $b_\pi = \text{NewTopRowConnect}(d_\pi, n)$
  $f_\pi = \text{AttachJellyfish}(b_\pi, n)$
  $O_{S_n}((I_\pi, J_\pi)) = \text{SymmGrpOrbit}(\pi, n)$
  for all $(I, J) \in O_{S_n}((I_\pi, J_\pi))$ **do**
    if $f_\pi(e_{(I,J)}) = +1$ **then**
      $O_{A_n}((I^+_\pi, J^+_\pi)).\text{Append}((I, J))$
    else
      $O_{A_n}((I^-_\pi, J^-_\pi)).\text{Append}((I, J))$
    end if
  end for
  $X^+_\pi = \sum_{(I,J)\in O_{A_n}((I^+_\pi,J^+_\pi))} E_{I,J}$
  $X^-_\pi = \sum_{(I,J)\in O_{A_n}((I^-_\pi,J^-_\pi))} E_{I,J}$
  $M = M + \lambda^+_\pi X^+_\pi + \lambda^-_\pi X^-_\pi$
end for
Return: M

We give an example that explicitly shows how to construct such a weight matrix in the Technical Appendix, in the case where $n = 2$, $k = 2$ and $l = 1$.

We appreciate that there will be some technical challenges when implementing Algorithm 1 given the current state of computer hardware. We discuss this in more detail in the Technical Appendix.

It is possible to extend our results by looking at linear layer functions that are equivariant to a direct product of alternating groups; this is given in full in the Technical Appendix.

7. Adding Features and Biases

7.1. Features

We have assumed throughout that the feature dimension for all of the layers appearing in the neural network is one. We can adapt all of the results that have been shown for the case where the feature dimension of the layers is greater than one.

Suppose that an r-order tensor has a feature space of dimension d_r. We now wish to find a basis for

$$\text{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k} \otimes \mathbb{R}^{d_k}, (\mathbb{R}^n)^{\otimes l} \otimes \mathbb{R}^{d_l}) \tag{45}$$

in the standard basis of $\mathbb{R}^n$.

Such a basis can be found by making the following substitutions, where now $i \in [d_l]$ and $j \in [d_k]$:

- replace $E_{I,J}$ by $E_{I,i,J,j}$ in (22), (41), and (42)
- relabel X_π by $X_{\pi,i,j}$, X^+_π by $X^+_{\pi,i,j}$, and X^-_π by $X^-_{\pi,i,j}$.

Consequently, a basis for (45) in the standard basis of $\mathbb{R}^n$ is given by the union of the two sets

$$\{X_{\pi,i,j} \mid \pi \in \Pi_{l+k,n-2}, i \in [d_l], j \in [d_k]\} \tag{46}$$

$$\left\{ X^+_{\pi,i,j}, X^-_{\pi,i,j} \;\middle|\; \begin{array}{l} \pi \in \Pi_{l+k,n} \setminus \Pi_{l+k,n-2}, \\ i \in [d_l], j \in [d_k] \end{array} \right\} \tag{47}$$

7.2. Biases

Including bias terms in the layer functions of a A_n-equivariant neural network is harder, but it can be done. For the learnable linear layers of the form $\text{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, Pearce–Crump (2022a) shows that the A_n-equivariance of the bias function, $\beta : ((\mathbb{R}^n)^{\otimes k}, \rho_k) \to ((\mathbb{R}^n)^{\otimes l}, \rho_l)$, needs to satisfy

$$c = \rho_l(g)c \tag{48}$$

for all $g \in A_n$ and $c \in (\mathbb{R}^n)^{\otimes l}$.

Since any $c \in (\mathbb{R}^n)^{\otimes l}$ satisfying (48) can be viewed as an element of $\text{Hom}_{A_n}(\mathbb{R}, (\mathbb{R}^n)^{\otimes l})$, to find the matrix form of c, all we need to do is to find a basis for $\text{Hom}_{A_n}(\mathbb{R}, (\mathbb{R}^n)^{\otimes l})$.

But this is simply a matter of applying Theorem 6.16, setting $k = 0$.

8. Related Work

The theory for the alternating group has its roots in the theory for the symmetric group and its links to the partition algebra. Jones (1994) constructed a surjective algebra homomorphism between the partition algebra $P_k(n)$ and the centraliser algebra of the symmetric group, $\text{End}_{S_n}((\mathbb{R}^n)^{\otimes k})$. Most notably, Benkart and Halverson went on to develop much of the theory for the duality between the symmetric group and the partition algebra in a number of important papers (2017; 2019a; 2019b). Bloss (2005) used the result of Jones (1994) to study the centralizer algebra of the alternating group, $\text{End}_{A_n}((\mathbb{R}^n)^{\otimes k})$. He showed that the partition algebra $P_k(n)$, which has a basis consisting of set partition diagrams having two rows of k vertices, is isomorphic to the centralizer algebra when $n \geq 2k + 2$. He also highlighted the difficulty of finding a diagrammatic approach for characterising $\text{End}_{A_n}((\mathbb{R}^n)^{\otimes k})$ in the remaining cases

since he recognised that the S_n orbits that correspond bijectively with the set partition diagrams split in these cases. Comes (2020) solved this problem and extended it to all Hom–spaces $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$, developing much of the theory in the process. In particular, he came up with the idea of using the determinant map to show how the S_n orbits split, and introduced jellyfish to represent this map in its diagrammatic form.

Finding the form of neural networks that are equivariant to a particular group has become an important area of research. Zaheer et al. (2017) introduced the first permutation equivariant neural network, called Deep Sets, for learning from sets in a permutation equivariant manner. Maron et al. (2019a) were the first to study the problem of classifying all of the linear permutation equivariant and invariant neural network layers, with their motivation coming from learning relations between the nodes of graphs. They characterised all of the learnable, linear, permutation equivariant layer functions in $\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ in the practical cases (specifically, when $n \geq k + l$). They used some equalities involving Kronecker products to find a number of fixed point equations which they solved to find a basis, in tensor form, for the layer functions under consideration.

As discussed in the Introduction, the approach taken in this paper to characterise all of the learnable, linear, equivariant layer functions in $\mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ is similar to the one seen in the papers written by Pearce–Crump (2022a; 2022b). They used various sets of set partition diagrams to characterise all of the learnable, linear, equivariant layer functions in $\mathrm{Hom}_G((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ when G is any of the following groups: the symmetric group S_n, the orthogonal group $O(n)$, the symplectic group $Sp(n)$, and the special orthogonal group $SO(n)$.

9. Conclusion

We are the first to show how the combinatorics underlying set partition diagrams, together with some jellyfish representing the determinant map, provides the theoretical background for constructing neural networks that are equivariant to the alternating group when the layers are some tensor power of $\mathbb{R}^n$. We looked at the problem of calculating the form of the learnable, linear, A_n–equivariant layer functions between such tensor power spaces in the standard basis of $\mathbb{R}^n$. We achieved this by finding a basis for the Hom–spaces in which these layer functions live. In particular, we showed how set partition diagrams correspond to a number of basis elements, and we used jellyfish to identify the individual basis elements when a set partition diagram corresponded to more than one basis element. In doing so, we calculated the number of parameters that appear in these layer functions. We also generalised our approach to show how to construct neural networks that are equivariant to local symmetries.

Acknowledgements

The author would like to thank his PhD supervisor Professor William J. Knottenbelt for being generous with his time throughout the author's period of research prior to the publication of this paper.

This work was funded by the Doctoral Scholarship for Applied Research which was awarded to the author under Imperial College London's Department of Computing Applied Research scheme. This work will form part of the author's PhD thesis at Imperial College London.

References

Barcelo, H. and Ram, A. Combinatorial Representation Theory, 1997. `arXiv:math/9707221`.

Benkart, G. and Halverson, T. Partition algebras $P_k(n)$ with $2k > n$ and the fundamental theorems of invariant theory for the symmetric group S_n. *J. London Math. Soc*, 99(2): 194–224, 2019a. URL `https://doi.org/10.1112/jlms.12175`.

Benkart, G. and Halverson, T. Partition Algebras and the Invariant Theory of the Symmetric Group. In Barcelo, H., Karaali, G., and Orellana, R. (eds.), *Recent Trends in Algebraic Combinatorics*, volume 16 of *Association for Women in Mathematics Series*, pp. 1–41. Springer, 2019b. URL `https://doi.org/10.1007/978-3-030-05141-9`.

Benkart, G., Halverson, T., and Harman, N. Dimensions of irreducible modules for partition algebras and tensor power multiplicities for symmetric and alternating groups. *J. Algebraic Combin.*, 46(1):77–108, 2017. ISSN 0925-9899. URL `https://doi.org/10.1007/s10801-017-0748-4`.

Bloss, M. The Partition Algebra as a Centralizer Algebra of the Alternating Group. *Communications in Algebra*, 33:2219–2229, 2005. URL `https://doi.org/10.1081/AGB-200063579`.

Brauer, R. On Algebras Which Are Connected with the Semisimple Continuous Groups. *Ann. Math.*, 38:857–872, 1937. ISSN 0003486X. URL `https://doi.org/10.2307/1968843`.

Ceccherini-Silberstein, T., Scarabotti, F., and Tolli, F. *Representation Theory of the Symmetric Groups*. Cambridge University Press, 2010.

Cohen, T. and Welling, M. Group Equivariant Convolutional Networks. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York,

USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Cohen, T., Weiler, M., Kicanaoglu, B., and Welling, M. Gauge Equivariant Convolutional Networks and the Icosahedral CNN. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/cohen19d.html.

Comes, J. Jellyfish Partition Categories. *Algebr Represent Theor*, 23:327–347, 2020. URL https://doi.org/10.1007/s10468-018-09851-7.

Goodman, R. and Wallach, N. R. *Symmetry, Representations and Invariants*. Springer, 2009.

Halverson, T. and Ram, A. Partition algebras. *European J. Combin.*, 26(6):869–921, 2005. ISSN 0195-6698. URL https://doi.org/10.1016/j.ejc.2004.06.005.

Halverson, T. and Ram, A. Gems from the Work of Georgia Benkart. *Notices of the American Mathematical Society*, 69(3):375–384, March 2022. URL https://doi.org/10.1090/noti2447.

Jones, V. The Potts model and the symmetric group. In Araki, H., Kawahigashi, Y., and Kosaki, H. (eds.), *Subfactors: Proceedings of the Taniguchi Symposium on Operator Algebras (Kyuzeso, 1993)*, pp. 259–267. World Scientific, 1994.

Kicki, P., Ozay, M., and Skrzypczynski, P. A Computationally Efficient Neural Network Invariant to the Action of Symmetry Subgroups, 2020. arXiv:2002.07528.

Kondor, R., Lin, Z., and Trivedi, S. Clebsch–Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/a3fc981af450752046be179185ebc8b5-Paper.pdf.

Lim, L.-H. and Nelson, B. J. What is an equivariant neural network?, 2022. arXiv:2205.07362.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. Invariant and Equivariant Graph Networks. In *International Conference on Learning Representations*, 2019a. URL https://openreview.net/forum?id=Syx72jC9tm.

Maron, H., Fetaya, E., Segol, N., and Lipman, Y. On the Universality of Invariant Networks. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4363–4371. PMLR, 09–15 Jun 2019b. URL https://proceedings.mlr.press/v97/maron19a.html.

Maron, H., Litany, O., Chechik, G., and Fetaya, E. On Learning Sets of Symmetric Elements. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6734–6744. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/maron20a.html.

Pearce-Crump, E. Connecting Permutation Equivariant Neural Networks and Partition Diagrams. arXiv:2212.08648, 2022a.

Pearce-Crump, E. Brauer's Group Equivariant Neural Networks. arXiv:2212.08630, 2022b.

Ravanbakhsh, S. Universal Equivariant Multilayer Perceptrons. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7996–8006. PMLR, 2020. URL http://proceedings.mlr.press/v119/ravanbakhsh20a.html.

Sagan, B. E. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer, 2000.

Segal, E. Group Representation Theory. Course Notes for Imperial College London, 2014.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. Deep Sets. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

Zhang, C., Liwicki, S., Smith, W., and Cipolla, R. Orientation-Aware Semantic Segmentation on Icosahedron Spheres. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3532–3540, Los Alamitos, CA, USA, nov 2019. IEEE Computer Society. doi: 10.1109/ICCV.2019.00363. URL https://doi.ieeecomputersociety.org/10.1109/ICCV.2019.00363.

A. The Weight Matrix for S_2 and A_2–Equivariant Linear Layer Functions from $(\mathbb{R}^2)^{\otimes 2}$ to $\mathbb{R}^2$

We first show how to find a basis of $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$, where, in this case, $n = 2$, $k = 2$, and $l = 1$.

To do this, we first need to find all set partitions in $\Pi_{1+2,2}$, that is, all set partitions π of 3 having at most 2 blocks. The first column of Figure 3 shows these set partitions in their equivalent diagram form, d_π. Each of these set partition diagrams corresponds to a basis element in $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$, by Theorem 5.1.

To obtain the basis elements themselves, we first recall that each set partition π corresponds to an orbit $O_{S_2}((I_\pi, J_\pi))$ coming from the action of S_2 on $[2]^{1+2}$. A representative of each orbit, called the block labelling, is given in the third column of Figure 3, and it is found by letting the ith position in $[1+2]$ be the label of the block in π containing the number i, where the blocks of π have themselves been labelled by letting B_1 be the block that contains the number $1 \in [1+2]$, and, if π has 2 blocks, letting B_2 be the other block in π.

Finally, we form a basis element X_π of $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ by summing over all matrix units in $\mathrm{Hom}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ that are indexed by the elements of $O_{S_2}((I_\pi, J_\pi))$, as stated in (22). These basis elements are given in the fourth column of Figure 3.

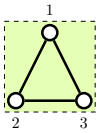
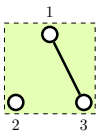
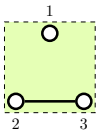
Set Partition Diagram d_π	Partition π	Block Labelling $(I_\pi \mid J_\pi)$	Standard Basis Element X_π
	$\{1,2,3\}$	$\{1 \mid 1,1\}$	$\begin{array}{c c} & \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$
	$\{1,2 \mid 3\}$	$\{1 \mid 1,2\}$	$\begin{array}{c c} & \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \end{array}$
	$\{1,3 \mid 2\}$	$\{1 \mid 2,1\}$	$\begin{array}{c c} & \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc\|cc} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right] \end{array}$
	$\{1 \mid 2,3\}$	$\{1 \mid 2,2\}$	$\begin{array}{c c} & \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{array}\right] \end{array}$

Figure 3. We use Theorem 5.2 to obtain a basis of $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ from all of the set partitions in $\Pi_{1+2,2}$.

Consequently, the weight matrix for an S_2–equivariant linear layer function from $(\mathbb{R}^2)^{\otimes 2}$ to $\mathbb{R}^2$ is of the form

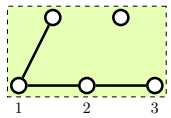
$$\begin{array}{c c} & \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc\|cc} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ \lambda_4 & \lambda_3 & \lambda_2 & \lambda_1 \end{array}\right] \end{array} \tag{49}$$

for scalars $\lambda_1, \dots, \lambda_4 \in \mathbb{R}$.

Next, we show how to find a basis of $\mathrm{Hom}_{A_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$. Again, we consider all set partitions in $\Pi_{1+2,2}$. Since $n = 2$, the S_2 orbit corresponding to each set partition in $\Pi_{1+2,2}$ splits because each set partition has either 1 or 2 blocks.

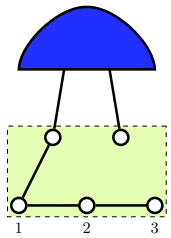
We show in full how to find the two basis elements X_π^+, X_π^- in $\mathrm{Hom}_{A_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ that corresponds to the first set partition diagram d_π in Figure 3, where $\pi = \{1,2,3\}$, and state what they are for the other set partition diagrams in Figure 4.

First, we flatten the set partition diagram d_π. Then we add a new top row consisting of $n = 2$ vertices, and connect the lowest numbered vertex in each block i of π to vertex i in the top row. Hence, we obtain the diagram b_π, which is



(50)

Next, we attach a two-legged jellyfish to the top row of vertices, giving



(51)

This is the diagram that is associated with the function f_π that is given in Proposition 6.13.

To obtain the two A_2 orbits $O_{A_2}((I_\pi^+, J_\pi^+))$ and $O_{A_2}((I_\pi^-, J_\pi^-))$ corresponding to π, we apply Theorem 6.14.

From Figure 3, we see that $O_{S_2}((I_\pi, J_\pi)) = \{(1,1,1),(2,2,2)\}$, and so, we have that

$$f_\pi(e_{(1,1,1)}) = \det \circ g_\pi(e_{(1,1,1)}) = \det(e_{(1,1)} + e_{(1,2)}) = +1 \tag{52}$$

and

$$f_\pi(e_{(2,2,2)}) = \det \circ g_\pi(e_{(2,2,2)}) = \det(e_{(2,1)} + e_{(2,2)}) = -1 \tag{53}$$

Hence, by (38) and (39), we have that

$$O_{A_2}((I_\pi^+, J_\pi^+)) = \{(1,1,1)\} \tag{54}$$

and

$$O_{A_2}((I_\pi^-, J_\pi^-)) = \{(2,2,2)\} \tag{55}$$

and so, by (41) and (42),

$$X_\pi^+ = E_{(1|1,1)} \tag{56}$$

and

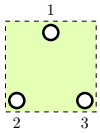
$$X_\pi^- = E_{(2|2,2)} \tag{57}$$

Consequently, from the matrices given in Figure 4, the weight matrix for an A_2–equivariant linear layer function from $(\mathbb{R}^2)^{\otimes 2}$ to $\mathbb{R}^2$ is of the form

$$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \left[\begin{array}{cc|cc} \lambda_1 & \lambda_3 & \lambda_5 & \lambda_7 \\ \lambda_8 & \lambda_6 & \lambda_4 & \lambda_2 \end{array}\right] \end{array} \tag{58}$$

for scalars $\lambda_1, \dots, \lambda_8 \in \mathbb{R}$.

Note that if $n = 3$, then we also need to consider the set partition diagram



(59)

corresponding to the set partition $\{1 \mid 2 \mid 3\}$, since this is a valid set partition in $\Pi_{1+2,3}$. In this case, only the set partition $\{1,2,3\}$ does not split. Hence, while the dimension of $\mathrm{Hom}_{S_3}((\mathbb{R}^3)^{\otimes 2}, \mathbb{R}^3)$ is 5, the dimension of $\mathrm{Hom}_{A_3}((\mathbb{R}^3)^{\otimes 2}, \mathbb{R}^3)$ is 9.

However, if $n \geq 4$, then none of the set partitions in $\Pi_{1+2,n}$ split, and so, in this case,

$$\mathrm{Hom}_{S_n}((\mathbb{R}^n)^{\otimes 2}, \mathbb{R}^n) = \mathrm{Hom}_{A_n}((\mathbb{R}^n)^{\otimes 2}, \mathbb{R}^n) \tag{60}$$

Consequently, the basis $\{X_\pi\}$, of size 5, is the same for each space.

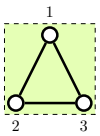
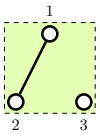
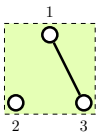
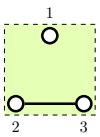
Set Partition Diagram d_π	Partition π	Standard Basis Element X_π^+	Standard Basis Element X_π^-
	$\{1,2,3\}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$
	$\{1,2 \mid 3\}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \end{array}$
	$\{1,3 \mid 2\}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right] \end{array}$
	$\{1 \mid 2,3\}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$	$\begin{array}{c\|c} & \begin{matrix}1,1 & 1,2 & 2,1 & 2,2\end{matrix} \\ \begin{matrix}1\\2\end{matrix} & \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \end{array}$

Figure 4. By considering which set partitions in $\Pi_{1+2,2}$ split, we obtain a basis of $\mathrm{Hom}_{A_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$.

B. Limitations and Feasibility

It is important to acknowledge that given the current limitations of hardware, there will be some challenges when implementing the neural networks that are discussed in this paper. In particular, significant engineering efforts will be needed to achieve the required scale because storing high-order tensors in memory is not a straightforward task. This was demonstrated by Kondor et al. (2018), who had to develop custom CUDA kernels in order to implement their tensor product based neural networks. Nevertheless, we anticipate that with the increasing availability of computing power, higher-order group equivariant neural networks will become more prevalent in practical applications. Notably, while the dimension of tensor power spaces increases exponentially with their order, the dimension of the space of equivariant maps between such tensor power spaces is often much smaller, and the corresponding matrices are typically sparse. Therefore, while storing these matrices may present some technical difficulties, it should be feasible with the current computing power that is available.

C. Equivariance to Local Symmetries

We can extend our results to looking at linear layer functions that are equivariant to a direct product of alternating groups; that is, we can construct neural networks that are equivariant to local symmetries. The case for an external tensor product of order 1 tensors can be found in (Maron et al., 2020); below, we show the equivariance for any tensors of any order.

We wish to find a basis for

$$\mathrm{Hom}_{A_{n_1}\times\cdots\times A_{n_p}}((\mathbb{R}^{n_1})^{\otimes k_1} \boxtimes \cdots \boxtimes (\mathbb{R}^{n_p})^{\otimes k_p}, (\mathbb{R}^{n_1})^{\otimes l_1} \boxtimes \cdots \boxtimes (\mathbb{R}^{n_p})^{\otimes l_p}) \tag{61}$$

where $\boxtimes$ is the external tensor product.

The Hom-space given in (61) is isomorphic to

$$\bigotimes_{r=1}^{p} \mathrm{Hom}_{A_{n_r}}((\mathbb{R}^{n_r})^{\otimes k_r}, (\mathbb{R}^{n_r})^{\otimes l_r}) \tag{62}$$

As Theorem 6.16 gives a basis for each individual Hom–space in (62), we can obtain a basis for the overall Hom–space (61) by forming all possible p–length Kronecker products of basis elements in the standard way for tensor product spaces.