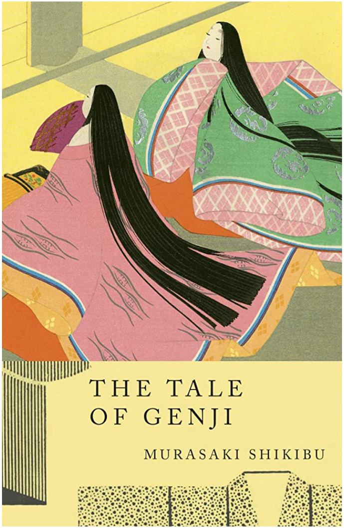
Imperial College London

Exploring Group Equivariant Neural Networks using Set Partition Diagrams

Online Machine Learning Seminar

Edward Pearce-Crump

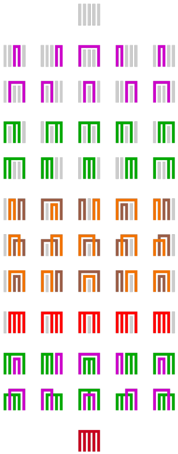
June 21, 2023



Jellyfish: https://unsplash.com/photos/v-ti3sccORY

The Tale of Genji: https://www.amazon.co.uk/Tale-Genji-Vintage-Classics/dp/0679729534

Chapters in the Tale of Genji: Set Partitions



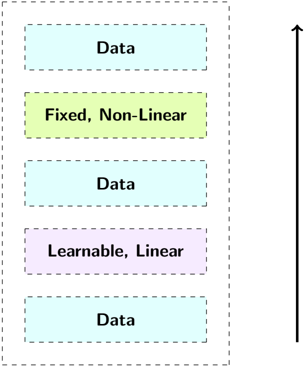
Attribution: T. Piesk https://commons.wikimedia.org/wiki/File:Set_partitions_5;_list;_Genji_symbols.svg

Outline

1. Review of Terminology

Neural Networks

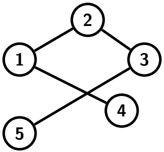
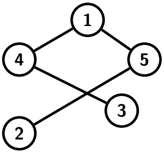
Neural networks consist of a **composition** of **layer functions**, where a layer function has the following typical form:



Shape of the Data

Physical processes often generate **data** that is high dimensional, and it can typically be represented in the form of a **high order tensor** (an element of $(\mathbb{R}^n)^{\otimes k}$) so that complex relationships can be captured between different features in the data.

Example: The adjacency matrix of a graph is a 2-order tensor

<table>
<tr><th>Graph</th><th>Matrix</th><th>Graph</th><th>Matrix</th></tr>
<tr>
<td>
</td>
<td>$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$</td>
<td>
</td>
<td>$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$</td>
</tr>
</table>

Symmetries

The data often comes with a certain type of **symmetry** that is baked into the data itself: e.g

- **permutations** (labelling of a set, labelling of the vertices of a graph etc.)
- **rotations** (of the circle in the plane, of a sphere, etc.)
- **translations** (of an object in an image, etc.)

Groups: The Mathematics of Symmetries

Symmetries in mathematics are given by **groups**.

A group is a non-empty set G together with a binary operation $\bullet$ such that the following axioms are satisfied:

1. **Closure:** for all $g, h \in G$, $g \bullet h \in G$
2. **Associativity:** for all $g, h, k \in G$, $(g \bullet h) \bullet k = g \bullet (h \bullet k)$
3. **Identity:** there exists a unique element $e \in G$ such that, for all $g \in G$, $g \bullet e = g = e \bullet g$
4. **Inverses:** for all $g \in G$, there exists an element $h \in G$ such that $g \bullet h = e = h \bullet g$.

Examples of Groups

- the **symmetric** group S_n: permutations of $[n] := \{1, \ldots, n\}$
- the **alternating** group A_n: subgroup of S_n consisting of all even permutations
- the **general linear** group $GL(n)$: the group of all invertible transformations $\mathbb{R}^n \to \mathbb{R}^n$
- the **special linear** group $SL(n)$: the subgroup of $GL(n)$ consisting of all invertible transformations whose determinant is $+1$.
- the **orthogonal** group $O(n)$: if we choose the standard basis of $\mathbb{R}^n$, this is the subgroup of $GL(n)$ consisting of matrices A such that $A^\top A = I_n$
- the **special orthogonal** group $SO(n)$: $O(n) \cap SL(n)$
- the **symplectic** group $Sp(n)$, $n = 2m$: if we choose the symplectic basis of $\mathbb{R}^n$, this is the subgroup of $GL(n)$ consisting of matrices A such that $A^\top JA = J$, where J is the block diagonal matrix consisting of m blocks of the form $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

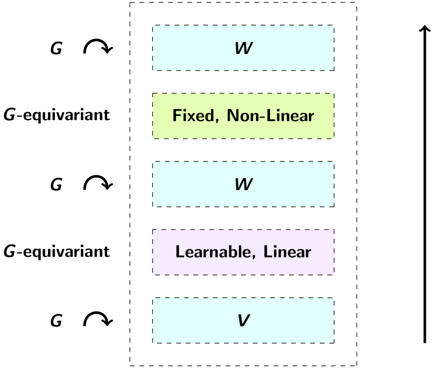
Group Actions and Group Representations

We are interested in **group actions**, where a group G acts on the elements of a set S, and, more specifically, if the elements of S index a basis of a vector space, then we are interested in **group representations**, which, in some sense, consider **groups as matrices**. This means that we can employ the tools of **linear algebra**.

Example: S_3 acts on $[3] = \{1, 2, 3\}$ by permutating its elements, and so, by indexing the standard basis $\{e_i\}$ of $\mathbb{R}^3$ by the elements of $[3]$, we obtain a representation $S_3 \to GL(\mathbb{R}^3)$ that is given by $\sigma(e_i) = e_{\sigma(i)}$, and is extended linearly on the basis.

Group Equivariant Neural Networks

These are neural networks where the data lives in a vector space that is a representation of a group, and the maps are **equivariant** to the group:



Group Equivariance: A Formal Definition

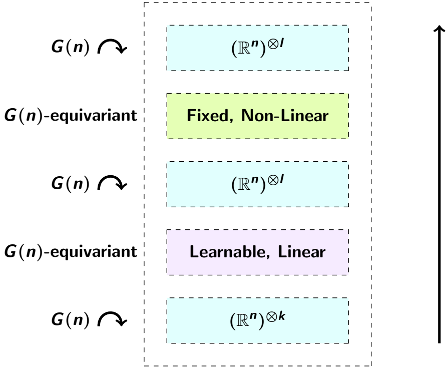
If $\rho_V : G \to GL(V)$ and $\rho_W : G \to GL(W)$ are two representations of G, then $\phi : V \to W$ is said to be G-**equivariant** if, for all $g \in G$ and $v \in V$,

$$\phi(\rho_V(g)[v]) = \rho_W(g)[\phi(v)] \tag{1}$$

The set of all *linear* G-equivariant maps between V and W is denoted by $\mathrm{Hom}_G(V, W)$, and, in particular, it forms a vector space.

2. Research Problem

Given the nature of the data that we wish to learn from, we are interested in group equivariant neural networks of the form:



$G(n)$ is a subgroup of $GL(n)$, and $(\mathbb{R}^n)^{\otimes k}$ is a representation of $G(n)$ given by the diagonal action over the tensor product:

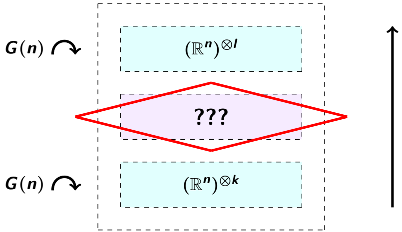
$$\rho_k(g)(v_1 \otimes \cdots \otimes v_k) := gv_1 \otimes \cdots \otimes gv_k \tag{2}$$

for all $g \in G(n)$ and for all vectors $v_i \in \mathbb{R}^n$.

Research Question

For different groups $G(n)$, can we characterise all of the possible $G(n)$–equivariant, learnable, linear layers that appear in a $G(n)$-equivariant neural network where the layers are some tensor power of $\mathbb{R}^n$?

In particular, can we find a **basis** or a **spanning set** of $\mathrm{Hom}_{G(n)}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ **in the standard basis of** $\mathbb{R}^n$?



Potential Benefits

1. **Less training data** is required: typically, the data does not need to be augmented.
2. These architectures come with **high levels of parameter sharing**: hence, there are **fewer parameters overall**.
3. **Reduction in time, effort and cost** needed to search for a neural network architecture: the form of the architectures is restricted by the symmetry group itself.
4. (Crucially, as we will see) we **do not need** to **decompose** the representation spaces $(\mathbb{R}^n)^{\otimes k}$ into **irreducibles** of $G(n)$; hence there is no need for change of basis transformations into the Fourier domain.

3. Relevant Literature

Most Relevant Literature

1. Zaheer et al. (2017), **arXiv:1703.06114**: introduced the first permutation equivariant neural network, called Deep Sets, for learning from sets in a permutation equivariant manner.
2. Maron et al. (2019), **arXiv:1812.09902**: characterised all of the learnable, linear, equivariant layer functions when the layers are some tensor power of $\mathbb{R}^n$ for the symmetric group S_n in the practical cases, by looking at fixed point equations representing the symmetric subspace.
3. Finzi et al. (2021), **arXiv:2104.09459**: constructed a numerical algorithm to find a basis to characterise the learnable, linear, equivariant layer functions when the layers are some tensor power of $\mathbb{R}^n$ for the orthogonal group $O(n)$, special orthogonal group $SO(n)$, and symplectic group $Sp(n)$, but only for small values of n and for small orders of the tensors, since their algorithm runs out of memory on higher values.

4. Results

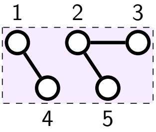
a) Symmetric Group S_n

We showed that there exists a bijective correspondence between

A Basis of $Hom_{S_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$

The Set Partitions of $[l + k]$ having at most n blocks

The S_n orbits of $[n]^{l+k}$



Pearce-Crump (2022): Connecting Permutation Equivariant Neural Networks and Partition Diagrams, **arXiv:2212.08648**, under review.

Bijection between S_n orbits and Set Partitions

Let (I, J) be a class representative of an S_n orbit of $[n]^{l+k}$, where $I \in [n]^l$ and $j \in [n]^k$. Then, writing

$$(I, J) = (i_1, i_2, \ldots, i_l, i_{l+1}, i_{l+2}, \ldots i_{l+k}) \tag{3}$$

we define the bijection, for all x, $y \in [l + k]$, by

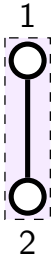
$$i_x = i_y \iff x, y \text{ are in the same block of } \pi \tag{4}$$

The bijection (4) is independent of the choice of class representative since

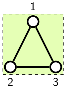
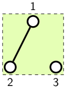
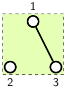
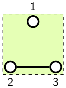
$$i_x = i_y \iff \sigma(i_x) = \sigma(i_y) \text{ for all } \sigma \in S_n \tag{5}$$

Notice that the LHS of (4) 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.

Example 1: $\mathrm{End}_{S_4}(\mathbb{R}^4)$ $(n = 4, k = 1, l = 1)$

Set Partition Diagram	Partition π	Block Labelling $(I_\pi \mid J_\pi)$	Standard Basis Element X_π
	$\{1, 2\}$	$\{1 \mid 1\}$	$\begin{array}{c c} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$
	$\{1 \mid 2\}$	$\{1 \mid 2\}$	$\begin{array}{c c} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$

Example 2: $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ $(n = 2, k = 2, l = 1)$

Set Partition Diagram	Partition π	Block Labelling $(I_\pi \mid J_\pi)$	Standard Basis Element X_π
	$\{1, 2, 3\}$	$\{1 \mid 1, 1\}$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc:cc} 1,1 & 1,2 & 2,1 & 2,2 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}$
	$\{1, 2 \mid 3\}$	$\{1 \mid 1, 2\}$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc:cc} 1,1 & 1,2 & 2,1 & 2,2 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}$
	$\{1, 3 \mid 2\}$	$\{1 \mid 2, 1\}$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc:cc} 1,1 & 1,2 & 2,1 & 2,2 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}$
	$\{1 \mid 2, 3\}$	$\{1 \mid 2, 2\}$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc:cc} 1,1 & 1,2 & 2,1 & 2,2 \\ \hline 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{array}$

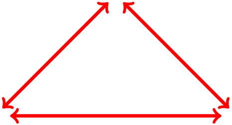
Other Relevant Literature

1. Martin (1990, 1994, 1996): first introduced the partition algebra $P_k(n)$ upon which these results are based.
2. Jones (1994): developed a surjective homomorphism between the partition algebra and the centraliser algebra on a k-order tensor of $\mathbb{R}^n$.
3. Benkart and Halverson (2019), **arXiv:1709.07751**: showed how the partition algebra can be used to construct the invariant theory of the symmetric group.

b) Orthogonal Group $O(n)$, Special Orthogonal Group $SO(n)$, Symplectic Group $Sp(n)$

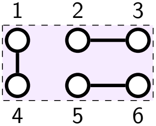
We showed that, for $O(n)$ and $Sp(n)$, there exists a bijective correspondence between

A Spanning Set of $\boldsymbol{Hom_{G(n)}}((\mathbb{R}^{\boldsymbol{n}})^{\otimes \boldsymbol{k}}, (\mathbb{R}^{\boldsymbol{n}})^{\otimes \boldsymbol{l}})$



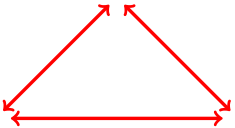
The Set Partitions of $[\boldsymbol{l} + \boldsymbol{k}]$ whose blocks come in pairs

First Fundamental Theorem for $\boldsymbol{G(n)}$:
A Spanning Set of $\boldsymbol{Hom_{G(n)}}((\mathbb{R}^{\boldsymbol{n}})^{\otimes \boldsymbol{l}+\boldsymbol{k}}, \mathbb{R})$



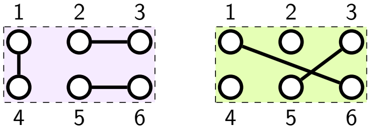
and, for $SO(n)$, there exists a bijective correspondence between

A Spanning Set of $Hom_{SO(n)}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$



The Set Partitions of $[l + k]$ whose blocks come in pairs together with those that have n elements removed and the rest come in pairs

First Fundamental Theorem for $SO(n)$: A Spanning Set of $Hom_{SO(n)}((\mathbb{R}^n)^{\otimes l+k}, \mathbb{R})$



Pearce-Crump (2022): Brauer's Group Equivariant Neural Networks, **arXiv:2212.08630**, ICML 2023 (live oral presentation).

Brauer's Invariant Argument

It can be shown that

$$C = \sum_{I \in [n]^l, J \in [n]^k} C_{I,J} E_{I,J} \tag{6}$$

is an element of $\mathrm{Hom}_{G(n)}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ (having chosen some basis of $\mathbb{R}^n$) if and only if the function

$$(\mathbb{R}^n)^{\otimes l+k} \to \mathbb{R} \tag{7}$$

which maps an element of the form

$$u(1) \otimes u(2) \otimes \cdots \otimes u(l) \otimes v(1) \otimes v(2) \otimes \cdots \otimes v(k) \tag{8}$$

to

$$\sum_{I \in [n]^l, J \in [n]^k} C_{I,J} \prod_{t=1}^{l} u_{i_t}(t) \prod_{r=1}^{k} v_{j_r}(r) \tag{9}$$

is an invariant for the group $G(n)$.

First Fundamental Theorem for $O(n)$

Suppose that $\mathbb{R}^n$ has associated with it a non-degenerate, symmetric, bilinear form $(\cdot,\cdot)$.

Pick the standard basis for $\mathbb{R}^n$, so that $(\cdot,\cdot)$ becomes the Euclidean inner product on $\mathbb{R}^n$.

If $f : (\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ is a polynomial function on elements in $(\mathbb{R}^n)^{\otimes(l+k)}$ of the form

$$u(1) \otimes u(2) \otimes \cdots \otimes u(l) \otimes v(1) \otimes v(2) \otimes \cdots \otimes v(k) \tag{10}$$

that is $O(n)$-invariant, then f must be a polynomial of the Euclidean inner products

$$(u(i), u(j)), (u(i), v(j)), (v(i), v(j)) \tag{11}$$

Hence, from Brauer's Invariant Argument, we get that

Theorem (Spanning Set of Invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ for $O(n)$)

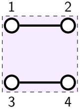
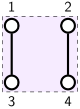
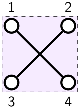
The functions

$$(z(1), z(2))(z(3), z(4)) \dots (z(l+k-1), z(l+k)) \tag{12}$$

where $z(1), \dots, z(l+k)$ *is a permutation of*

$$u(1), u(2), \dots, u(l), v(1), v(2), \dots, v(k)$$

form a spanning set of invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ *for* $O(n)$.

Set Partition Diagram	Inner Products	Spanning Set Element
	$(u(1), u(2))(v(1), v(2))$	$\begin{array}{c} \\ 1,1 \\ 1,2 \\ 2,1 \\ 2,2 \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] \end{array}$
	$(u(1), v(1))(u(2), v(2))$	$\begin{array}{c} \\ 1,1 \\ 1,2 \\ 2,1 \\ 2,2 \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$
	$(u(1), v(2))(u(2), v(1))$	$\begin{array}{c} \\ 1,1 \\ 1,2 \\ 2,1 \\ 2,2 \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,2 & 2,1 & 2,2 \end{array} \\ \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$

First Fundamental Theorem for $Sp(n)$

Suppose that $\mathbb{R}^n$ $(n = 2m)$ has associated with it a non-degenerate, skew-symmetric, bilinear form $\langle \cdot, \cdot \rangle$.

Choosing the **symplectic basis**

$$\widetilde{B} := \{e_1, e_{1'}, \ldots, e_m, e_{m'}\} \tag{13}$$

for $\mathbb{R}^n$, where the i^{th} basis vector in the set has a 1 in the i^{th} position and a 0 elsewhere, which satisfies the relations

$$\langle e_\alpha, e_\beta \rangle = \langle e_{\alpha'}, e_{\beta'} \rangle = 0 \tag{14}$$

$$\langle e_\alpha, e_{\beta'} \rangle = -\langle e_{\alpha'}, e_\beta \rangle = \delta_{\alpha,\beta} \tag{15}$$

we have that $\langle \cdot, \cdot \rangle$ becomes the skew product

$$\langle x, y \rangle = \sum_{r=1}^{m} (x_r y_{r'} - x_{r'} y_r) = \sum_{i,j} \langle e_i, e_j \rangle x_i y_j \tag{16}$$

for all $x, y \in \mathbb{R}^n$.

Note that, in this basis, the non-degenerate, *symmetric*, bilinear form $(\cdot,\cdot)$ which we can also associate with $\mathbb{R}^n$, becomes the Euclidean inner product on $\mathbb{R}^n$ since the symplectic basis is standard with respect to $(\cdot,\cdot)$.

Then, if $f : (\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ is a polynomial function on elements in $(\mathbb{R}^n)^{\otimes(l+k)}$ of the form

$$u(1) \otimes u(2) \otimes \cdots \otimes u(l) \otimes v(1) \otimes v(2) \otimes \cdots \otimes v(k) \tag{17}$$

that is $Sp(n)$-invariant, then f must be a polynomial of the Euclidean inner products

$$(u(i), v(j)) \tag{18}$$

together with the skew products

$$\langle u(i), u(j)\rangle, \langle v(i), v(j)\rangle \tag{19}$$

such that $i < j$ in (19).

Hence, from Brauer's Invariant Argument, we get that

Theorem (Spanning Set of Invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ for $Sp(n)$, $n = 2m$)

The functions

$$[z(1), z(2)][z(3), z(4)] \dots [z(l+k-1), z(l+k)] \tag{20}$$

where $z(1), \dots, z(l+k)$ *is a permutation of*

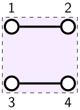
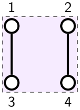
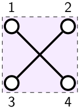
$$u(1), u(2), \dots, u(l), v(1), v(2), \dots, v(k)$$

and

$$[z(i), z(i+1)] := \begin{cases} (z(i), z(i+1)) & \text{if } z(i) = u(j) \text{ and } z(i+1) = v(m), \\ & \text{or } z(i) = v(m) \text{ and } z(i+1) = u(j), \\ & \text{for some } j \in [l],\ m \in [k] \\ \langle z(i), z(i+1) \rangle & \text{otherwise.} \end{cases} \tag{21}$$

form a spanning set of invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ *for* $Sp(n)$, *with* $n = 2m$.

Example 2: $\mathrm{End}_{Sp(2)}((\mathbb{R}^2)^{\otimes 2})$ $(n = 2, k = 2, l = 2)$

Set Partition Diagram	Inner/Skew Products	Spanning Set Element
	$\langle u(1), u(2)\rangle\langle v(1), v(2)\rangle$	$\begin{array}{c} \\ 1,1 \\ 1,1' \\ 1',1 \\ 1',1' \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,1' & 1',1 & 1',1' \end{array} \\ \left[\begin{array}{cc\|cc} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ \hline 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array}$
	$(u(1), v(1))(u(2), v(2))$	$\begin{array}{c} \\ 1,1 \\ 1,1' \\ 1',1 \\ 1',1' \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,1' & 1',1 & 1',1' \end{array} \\ \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$
	$(u(1), v(2))(u(2), v(1))$	$\begin{array}{c} \\ 1,1 \\ 1,1' \\ 1',1 \\ 1',1' \end{array} \begin{array}{c} \begin{array}{cccc} 1,1 & 1,1' & 1',1 & 1',1' \end{array} \\ \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$

First Fundamental Theorem for $SO(n)$

Suppose that $\mathbb{R}^n$ has associated with it a non-degenerate, symmetric, bilinear form $(\cdot,\cdot)$. Choose the standard basis for $\mathbb{R}^n$, so that $(\cdot,\cdot)$ becomes the Euclidean inner product on $\mathbb{R}^n$.

If $f : (\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ is a polynomial function on elements in $(\mathbb{R}^n)^{\otimes(l+k)}$ of the form

$$u(1) \otimes u(2) \otimes \cdots \otimes u(l) \otimes v(1) \otimes v(2) \otimes \cdots \otimes v(k) \tag{22}$$

that is $SO(n)$-invariant, then it must be a polynomial of the Euclidean inner products

$$(u(i), u(j)), (u(i), v(j)), (v(i), v(j)) \tag{23}$$

together with the $n \times n$ subdeterminants of the $n \times (l+k)$ matrix M having as its columns:

$$M := \begin{pmatrix} | & | & & | & | & | & & | \\ u(1) & u(2) & \dots & u(l) & v(1) & v(2) & \dots & v(k) \\ | & | & & | & | & | & & | \end{pmatrix} \tag{24}$$

Hence, from Brauer's Invariant Argument, we get that

Theorem (Spanning Set of Invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ for $SO(n)$)

Functions of the form

$$(z(1), z(2))(z(3), z(4)) \dots (z(l+k-1), z(l+k)) \tag{25}$$

together with functions of the form

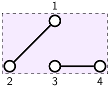
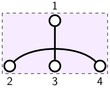
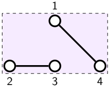
$$\det(z(1), \dots, z(n))(z(n+1), z(n+2)) \dots (z(l+k-1), z(l+k)) \tag{26}$$

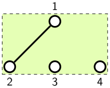
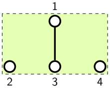
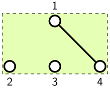
where $z(1), \dots, z(l+k)$ is a permutation of

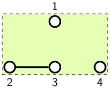
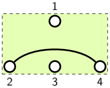
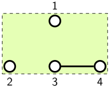
$$u(1), u(2), \dots, u(l), v(1), v(2), \dots, v(k)$$

form a spanning set of invariants $(\mathbb{R}^n)^{\otimes(l+k)} \to \mathbb{R}$ for $SO(n)$.

Example 3: $\mathrm{Hom}_{SO(2)}((\mathbb{R}^2)^{\otimes 3}, \mathbb{R}^2)$ $(n = 2, k = 3, l = 1)$

Set Partition Diagram	Inner Products	Spanning Set Element
	$(u(1), v(1))(v(2), v(3))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc\|cc\|cc\|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}$
	$(u(1), v(2))(v(1), v(3))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc\|cc\|cc\|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}$
	$(u(1), v(3))(v(1), v(2))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc\|cc\|cc\|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$

Set Partition Diagram	Inner Products	Spanning Set Element
	$\det(v(2), v(3))(u(1), v(1))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc|cc|cc|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{array}$
	$\det(v(1), v(3))(u(1), v(2))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc|cc|cc|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \end{array}$
	$\det(v(1), v(2))(u(1), v(3))$	$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{cc|cc|cc|cc} 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{array}$

Set Partition Diagram	Inner Products	Spanning Set Element
	$\det(u(1), v(3))(v(1), v(2))$	$\begin{array}{c\|cc\|cc\|cc\|cc} & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 2 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{array}$
	$\det(u(1), v(2))(v(1), v(3))$	$\begin{array}{c\|cc\|cc\|cc\|cc} & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 2 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{array}$
	$\det(u(1), v(1))(v(2), v(3))$	$\begin{array}{c\|cc\|cc\|cc\|cc} & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,1 & 2,1,2 & 2,2,1 & 2,2,2 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{array}$

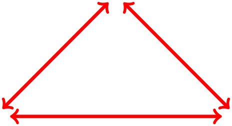
Other Relevant Literature

1. Brauer (1937): first introduced the Brauer algebra for the purpose of understanding the centraliser algebras of the groups $O(n)$, $SO(n)$ and $Sp(n)$
2. Grood (1999): investigated the representation theory of the Brauer–Grood algebra

c) Alternating Group A_n

We showed that there exists a bijective correspondence between

A Basis of $Hom_{A_n}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$



The Set Partitions of $[l+k]$
having at most $n-2$ blocks
correspond to 1 A_n orbit
and those having either $n-1$ or n blocks
correspond to 2 A_n orbits

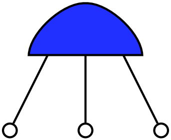
The A_n orbits of $[n]^{l+k}$

Pearce-Crump (2023): How Jellyfish Characterise Group Equivariant Neural Networks, **arXiv:2301.10152**, ICML 2023 (poster presentation).

The set partitions that correspond to more than one A_n orbit are said to **split**.

The basis elements corresponding to set partitions that do not split can be found in exactly the same way as for the symmetric group S_n.

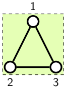
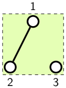
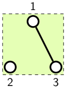
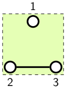
To find the basis elements corresponding to set partitions that split, we use n-legged **jellyfish**



which correspond to the **determinant map** $(\mathbb{R}^n)^{\otimes n} \to \mathbb{R}$ defined on the standard basis by

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

From before: $\mathrm{Hom}_{S_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ $(n = 2, k = 2, l = 1)$

Set Partition Diagram	Partition π	Block Labelling $(I_\pi \mid J_\pi)$	Standard Basis Element X_π
	$\{1, 2, 3\}$	$\{1 \mid 1, 1\}$	$\begin{matrix} & 1,1 & 1,2 & 2,1 & 2,2 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{matrix}$
	$\{1, 2 \mid 3\}$	$\{1 \mid 1, 2\}$	$\begin{matrix} & 1,1 & 1,2 & 2,1 & 2,2 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \end{matrix}$
	$\{1, 3 \mid 2\}$	$\{1 \mid 2, 1\}$	$\begin{matrix} & 1,1 & 1,2 & 2,1 & 2,2 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 \end{matrix}$
	$\{1 \mid 2, 3\}$	$\{1 \mid 2, 2\}$	$\begin{matrix} & 1,1 & 1,2 & 2,1 & 2,2 \\ 1 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 \end{matrix}$

What about $\mathrm{Hom}_{A_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$?

Procedure

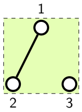
We follow the **procedure** given below (for **general** n, k, l):

1. Check which of the (k, l)–set partition diagrams appearing in the S_n case split and which do not. In this case, all four $(2, 1)$–set partition diagrams split.

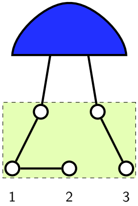
2. For those that do not split, the basis matrix is the same as for the S_n case.

3. Otherwise, we construct a **jellyfish diagram** for each (k, l)–set partition diagram that **splits**, as follows: first, we flatten it, maintaining the order of the vertices, then we add a new top row of n vertices and connect the lowest numbered vertex in each block i to vertex i in the top row, and finally we attach an n-legged jellyfish to the top row of vertices.

For example, for the second set partition diagram



we obtain the following jellyfish diagram:



We can show that the jellyfish diagram corresponds to a **map** in $\text{Hom}_{A_n}((\mathbb{R}^n)^{\otimes l+k}, \mathbb{R})$ that sends standard basis vectors indexed by the elements of the S_n orbit that corresponds to the original (k, l)–set partition diagram to ± 1, and to 0 otherwise.

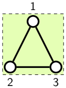
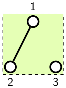
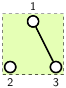
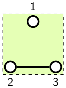
4. We then calculate the **two A_n orbits** that the S_n orbit **splits into**, using the **possible outcomes** of the **map** on standard basis vectors indexed by its elements, namely ± 1, as **separate classes**.

For the jellyfish diagram given above, as the S_2 orbit is $\{(1,1,2),(2,2,1)\}$, the map takes $e_{(1,1,2)}$ to $+1$ and $e_{(2,2,1)}$ to -1. Hence the S_2 orbit splits into two A_2 orbits, namely $\{(1,1,2)\}$ and $\{(2,2,1)\}$.

5. Finally, we obtain the two basis matrices X^+ and X^-, each of which is a sum of the matrix units in $\mathrm{Hom}((\mathbb{R}^n)^{\otimes k}, (\mathbb{R}^n)^{\otimes l})$ that are indexed by the elements of the A_n orbits.

For our example, we have that $X^+ = E_{(1|1,2)}$ and $X^- = E_{(2|2,1)}$

Example: $\mathrm{Hom}_{A_2}((\mathbb{R}^2)^{\otimes 2}, \mathbb{R}^2)$ $(n = 2, k = 2, l = 1)$

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

Other Relevant Literature

1. Bloss (2005): studied the centraliser algebra of the alternating group, adapting the result of Jones (1994) for that of the symmetric group.
2. Comes (2020), **arXiv:1612.05182**: largely determined the theory of alternating group A_n equivariance; however, they relied heavily on the language of category theory in their exposition.

Other Papers

1. Pearce-Crump (2023): Categorisation of Group Equivariant Neural Networks, **arXiv:2304.14144**, under review: developed a category theoretic framework around the characterisations given above, leading to:
2. Pearce-Crump (2023): An Algorithm for Computing with Brauer's Group Equivariant Neural Network Layers, **arXiv:2304.14165**, under review: allows us to compute with the layers characterised in Brauer's Group Equivariant Neural Networks in a faster way, using decompositions into Kronecker product matrices.

5. Closing Remarks

Limitations and Discussion

- Given the current **limitations of hardware**, tensor product neural networks require **significant engineering efforts** in order to achieve the required scale
- This is 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.