FAST COMPUTATION OF PERMUTATION EQUIVARIANT LAYERS WITH THE PARTITION ALGEBRA

Charles Godfrey,[1] **Michael G. Rawson,**[1] **Davis Brown,**[1] **and Henry Kvinge**[1,2]
[1]Pacific Northwest National Laboratory, [2] Department of Mathematics, University of Washington
{first}.{last}@pnnl.gov

ABSTRACT

Linear neural network layers that are either equivariant or invariant to permutations of their inputs form core building blocks of modern deep learning architectures. Examples include the layers of DeepSets, as well as linear layers occurring in attention blocks of transformers and some graph neural networks. The space of permutation equivariant linear layers can be identified as the invariant subspace of a certain symmetric group representation, and recent work parameterized this space by exhibiting a basis whose vectors are sums over orbits of standard basis elements with respect to the symmetric group action. A parameterization opens up the possibility of learning the weights of permutation equivariant linear layers via gradient descent. The space of permutation equivariant linear layers is a generalization of the partition algebra, an object first discovered in statistical physics with deep connections to the representation theory of the symmetric group, and the basis described above generalizes the so-called orbit basis of the partition algebra. We exhibit an alternative basis, generalizing the diagram basis of the partition algebra, with computational benefits stemming from the fact that the tensors making up the basis are low rank in the sense that they naturally factorize into Kronecker products. Just as multiplication by a rank one matrix is far less expensive than multiplication by an arbitrary matrix, multiplication with these low rank tensors is far less expensive than multiplication with elements of the orbit basis. Finally, we describe an algorithm implementing multiplication with these basis elements.

1 INTRODUCTION

Invariance or equivariance to application-driven symmetry groups has served as a guiding light for the design of neural network architectures for over two decades, dating back at least to the introduction of convolutional networks (Fukushima, 1980). In the case where the underlying symmetries are permutations, several families of architectures have appeared in the last five years: DeepSets (Zaheer et al., 2018) (and its successors), neural networks operating on graphs (where invariance to node permutations is a natural desiderata) (Maron et al., 2018), and transformers (Vaswani et al., 2017).

Permutation equivariant linear layers are linear maps $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ equivariant with respect to the symmetric group Σ_n acting "diagonally" on tensors on the domain and target (Section 3 includes background and more detailed definitions). These include linear layers used in DeepSets and some layers of transformers (for the later case see (Kim et al., 2021), (Bronstein et al., 2021, §5.4)) as special cases and as such are a fundamental building block for permutation equivariant/invariant models. The result of (Maron et al., 2018, Thm. 1), which computes a basis for the vector space parameterizing permutation equivariant linear layers, opens up the possibility of *learning* these layers, i.e. treating the parameters as weights and optimizing them with some form of gradient descent.

There are good reasons related to computational efficiency for avoiding the basis introduced in (Maron et al., 2018): under a natural identification, permutation equivariant linear layers can be identified as tensors living in $(\mathbb{R}^n)^{\otimes(m+m')}$. Just as multiplication by a rank one matrix of the form uv^T, where u and v are vectors, can be executed far faster with sequential dot products than multiplication with an arbitrary matrix, contraction with a tensor in $(\mathbb{R}^n)^{\otimes(m+m')}$ that decomposes as a Kronecker product, say $u \otimes v$ where $u \in (\mathbb{R}^n)^{\otimes p}, v \in (\mathbb{R}^n)^{\otimes p'}$ and $p + p' = m + m'$, will generally be less expensive than contraction with an arbitrary tensor in $(\mathbb{R}^n)^{\otimes(m+m')}$. Unfortunately, many of the basis vectors found in (Maron et al., 2018, Thm. 1) lack such a decomposition (an example is given in Appendix D).

Our main result, Theorem 5.4 below, exhibits an alternative basis for permutation equivariant linear layers in which all but one basis vector are explicitly constructed as Kronecker products. In this construction, we exploit the fact that the direct sum of the vector spaces of permutation equivariant linear layers, ranging from zero to infinity, forms an *algebra* in which multiplication is the Kronecker product.

In the case where $m = m'$, this basis is known: permutation equivariant layers $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m}$ are exactly the *partition algebra*, an object first discovered in the context of statistical mechanics in the mid 1990s that has since been extensively studied by mathematicians (Halverson & Ram, 2005; Benkart & Halverson, 2017) , and the basis of Theorem 5.4 is simply the "diagram basis" of the partition algebra (that of (Maron et al., 2018, Thm. 1) is known as the "orbit sum" basis). We view this connection as a beautiful example of an idea that originated in physics and made its way into deep learning architectures which are now being applied to model physical phenomena, for example predicting molecular geometry using the ZINC dataset (Sterling & Irwin, 2015).

For the sake of concreteness we work over $\mathbb{R}$ throughout, but all mathematical results in this paper apply over any field, including the complex numbers $\mathbb{C}$ and the finite field $\mathbb{Z}/2$. This is not generality for generality's sake, as the latter is used in quantized neural networks.

The contributions of this paper include: (i) an explicit construction of an alternative basis for the space of permutation equivariant linear layers, (ii) verification that in the case where the permutation equivariant linear layers are the partition algebra, our basis coincides with the diagram basis, and (iii) outlining an algorithm implementing multiplication with these basis elements, and in the process showing they recover the efficient operations of (Pan & Kondor, 2022).

2 Related Work

In Appendix C, we show that the basis of Theorem 5.4 recovers the operations described in (Pan & Kondor, 2022, §4-5). Note however that while their work does prove that the set of "sum/transfer-/broadcast" tensors, that they define, has the same cardinality as the basis of (Maron et al., 2018, Thm. 1), it does *not* demonstrate that those "sum/transfer/broadcast" tensors are linearly independent nor that they span the space of permutation equivariant linear layers.

By (Pan & Kondor, 2022, Appendix), the basis of Theorem 5.4 also coincides with the one used in (Maron et al., 2018, Appendix). However, the latter authors also omit a proof of linear independence/spanning. The dimension of the space of permutation-equivariant linear layers grows extremely rapidly with $m + m'$ and (for sufficiently large n) equals the Bell number $B(m + m')$, see (Maron et al., 2018, Thm. 1). Thus while the case of $m = m' = 2$ considered in their work may have been tractable as a one-off case, those of higher $m + m'$ will not be.

The partition algebra was first discovered (independently) in (Martin, 1991; Jones, 1994). Its structure (in terms of generators and relations) and connections with representation theory of the symmetric group was identified in (Halverson & Ram, 2005) (see also (Benkart & Halverson, 2017) for more recent developments). In the case $m = m'$, our Theorem 5.4 just gives an explicit description of the aforementioned diagram basis in terms of Kronecker products of diagonal tensors.

After completing this work, we became aware of the recent article (Pearce-Crump, 2023), which also points out the connection between permutation equivariant linear layers and partition algebras. This article has technical overlap with ours in Section 3 (which overlap with multiple references for that matter), however it does not include our main results (those of Sections 4 and 5), nor algorithms for multiplying with tensors representing permutation equivariant maps such as those of Appendix C, nor connections with the operations of (Pan & Kondor, 2022, §4-5).

3 Background

We fix a natural number $n \in \mathbb{N}$. In applications, this is the cardinality of the set on which a permutation equivariant/invariant model operates. The vector space $\mathbb{R}^n$ comes with a natural action of the symmetric group on n elements, denoted Σ_n, permuting the basis vectors $\{e_1, \dots, e_n\}$:

$$\sigma \cdot e_i = e_{\sigma(i)} \text{ for } \sigma \in \Sigma_n, i = 1, \dots, n. \tag{3.1}$$

For each natural number $m \in \mathbb{N}$ the m-th tensor power

$$(\mathbb{R}^n)^{\otimes m} = \underbrace{\mathbb{R}^n \otimes \mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n}_{m \text{ times}} \tag{3.2}$$

is also a representation of Σ_n with the "diagonal" action, defined for a $\sigma \in \Sigma_n$ and basis tensor $v_1 \otimes \cdots \otimes v_m \in (\mathbb{R}^n)^{\otimes m}$ as $\sigma \cdot (v_1 \otimes \cdots \otimes v_m) := (\sigma \cdot v_1) \otimes \cdots \otimes (\sigma \cdot v_m)$.

Remark 3.3. When $m = 0$ we adopt the convention $(\mathbb{R}^n)^{\otimes m} = \mathbb{R}$ with the trivial Σ_n action.

We denote the vector space of *all* (not necessarily equivariant) linear maps $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ by $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})$. A linear map $\varphi \in \mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})$ is Σ_n-equivariant if and only if $\varphi(\sigma \cdot v) = \sigma \cdot \varphi(v)$ for all $\sigma \in \Sigma_n, v \in (\mathbb{R}^n)^{\otimes m}$, and we denote the vector space of such Σ_n-equivariant linear maps $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$.

Problem 3.4. *Parameterize* $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$*, for example by giving a basis for it as a subspace of* $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})$.

The next two elementary results will be useful in what follows. Proofs are deferred to the appendix.

Lemma 3.5 (cf. (Halverson & Ram, 2005, §3))**.** $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ *is isomorphic to the* Σ_n*-invariant subspace of the tensor power* $(\mathbb{R}^n)^{\otimes (m+m')}$.

Corollary 3.6. *The dimension of* $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ *is a function of the* ***sum*** $m + m'$ *(for fixed* n*). Moreover, given any parametrization of the* Σ_n*-invariant subspace of* $(\mathbb{R}^n)^{\otimes l}$ *for some* $l \in \mathbb{N}$ *there is a simple recipe to produce parametrizations of* $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ *for all* $m + m' = l$.

Lemma 3.5 and Corollary 3.6 reduce Problem 3.4 to the computation of the invariant subspace $((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$. Now we give our notation for partitions. Given a tuple $(i_1, \dots, i_l) \in \{1, \dots, n\}^l$, define subsets $S(i_1, \dots, i_l)_j \subseteq \{1, \dots, l\}$ for $j = 1, \dots, n$ by $S(i_1, \dots, i_l)_j = \{k \in \{1, \dots, l\} \mid i_k = j\}$. By construction these subsets are pairwise disjoint and their union is $\{1, \dots, l\}$; thus, they form a *set partition of* $\{1, \dots, l\}$, which we will denote by

$$\Pi(i_1, \dots, i_l) = \{S(i_1, \dots, i_l)_j \mid j = 1, \dots, n\}. \tag{3.7}$$

In this way we obtain a map from tuples $(i_1, \dots, i_l) \in \{1, \dots, n\}^l$ to set partitions $\Pi(i_1, \dots, i_l)$ of $\{1, \dots, l\}$. The next lemma shows that the partition $\Pi(i_1, \dots, i_l)$ uniquely characterizes the Σ_n-orbit of $(i_1, \dots, i_l)$. For an illustration of the lemma in an explicit example, see Appendix D.

Lemma 3.8 (cf. (Jones, 1994, §1), (Benkart & Halverson, 2017, §5.2))**.** *The tuples* $(i_1, \dots, i_l)$ *and* $(i'_1, \dots, i'_l)$ *lie in the same* Σ_n *orbit if and only if they give rise to the same partition of* $\{1, \dots, l\}$*, i.e.* $\Pi(i_1, \dots, i_l) = \Pi(i'_1, \dots, i'_l)$.

The main theorem of Maron et al. (2018) solves Problem 3.4 by exhibiting a basis described as follows: Let $\mathcal{P}$ be a fixed set partition of of $\{1, \dots, l\}$, and define

$$e_{\mathcal{P}} = \sum_{(i_1, \dots, i_l) : \Pi(i_1, \dots, i_l) = \mathcal{P}} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_l} \in \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{3.9}$$

Theorem 3.10 ((Maron et al., 2018, Thm. 1))**.** *The vectors* $\{e_{\mathcal{P}} \mid \mathcal{P} \text{ is a set partition of } \{1, \dots, l\}\}$ *form a basis of* $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$.

In Appendix A we provide intuition behind this construction.

4 The partition algebra $T(\mathbb{R}^n)$ and its Σ_n-invariant subalgebra

The invariant subspaces $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ assemble to form an algebra. Multiplication in this algebra can be interpreted as an operation that produces "new permutation equivariant layers from old" by taking two permutation equivariant linear layers $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ and $(\mathbb{R}^n)^{\otimes p} \to (\mathbb{R}^n)^{\otimes p'}$ and forming a third, $(\mathbb{R}^n)^{\otimes (m+p)} \to (\mathbb{R}^n)^{\otimes (m'+p')}$, by taking the tensor product.

Lemma 4.1. *For any $p, p' \in \mathbb{N}$ such that $p + p' = l$, there is a natural bilinear map*

$$\left((\mathbb{R}^n)^{\otimes p}\right)^{\Sigma_n} \times \left((\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \to \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \text{ taking } (v, w) \mapsto v \otimes w. \tag{4.2}$$

The direct sum $\bigoplus_{l \in \mathbb{N}} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ forms an $\mathbb{R}$-algebra since the bilinear maps of Lemma 4.1 are associative and unital in a suitable sense.

5 A NEW-OLD BASIS FOR PERMUTATION EQUIVARIANT LINEAR LAYERS

For each partition $\mathcal{P}$ of $\{1, \ldots, l\}$, we let $e_{\mathcal{P}} \in \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ be the basis element associated by Lemmas 3.8 and A.2. We claim there is another way to generate basis elements from partitions, with potential computational advantages: given $\mathcal{P} = \{S_1, \ldots, S_n\}$, there is a multilinear map

$$\Phi_{\mathcal{P}} : \prod_{i=1}^{n} \Big((\bigotimes_{j=1}^{|S_i|} \mathbb{R}^n)^{\Sigma_n} \Big) \xrightarrow{\mu} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \xrightarrow{\tau} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \tag{5.1}$$

where μ is obtained from repeated use of the bilinear maps from Lemma 4.1 and τ is a permutation of $\{1, \ldots, l\}$ (acting on tensor factors) whose inverse arranges the sets S_i in successive contiguous blocks. [1] Within each $(\bigotimes_{j=1}^{|S_i|} \mathbb{R}^n)^{\Sigma_n}$ there is a basis element $e_{\{\{1,\ldots,|S_i|\}\}}$ corresponding to the partition with one set, i.e. a diagonal tensor. We define:

$$d_{\mathcal{P}} := \Phi_{\mathcal{P}}(e_{\{\{1,\ldots,|S_1|\}\}}, e_{\{\{1,\ldots,|S_2|\}\}}, \cdots, e_{\{\{1,\ldots,|S_n|\}\}}). \tag{5.2}$$

Remark 5.3. The choice of permutation of (labels of) $\{1, \ldots, l\}$ used to define $\Phi_{\mathcal{P}}$ is irrelevant, as any two choices differ by a sequence of permutations of the individual $S_1, \ldots, S_n$, and $e_{\{S_i\}}$ is invariant to permutations of S_i.

Theorem 5.4. *The vectors* $\left\{ d_{\mathcal{P}} \mid \mathcal{P} \text{ is a set partition of } \{1, \ldots, l\} \right\}$ *are a basis for* $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$.

As mentioned in the introduction, the space of permutation equivariant linear layers $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m}$ is a partition algebra.

Proposition 5.5. *In the $m = m'$ case, the basis for $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m})^{\Sigma_n}$ constructed in Theorem 5.4 coincides with the diagram basis of the partition algebra denoted by "$L_\sim$" in (Jones, 1994).*

By construction, the tensors of $d_{\mathcal{P}}$ are factored Kronecker products. In Appendix C, we describe an algorithm for computing the Σ_n-equivariant map $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ corresponding to multiplication with the tensor $d_{\mathcal{P}} \in ((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$. While we leave an analysis of computational cost (e.g. in terms of FLOPs) to future work, the algorithm shows the factorization of $d_{\mathcal{P}}$ makes multiplying with them computationally efficient (in particular more efficient than the the elements $e_{\mathcal{P}}$, see Appendix D). The description in Appendix C also shows that the $d_{\mathcal{P}}$ recover the "sum/transfer/broadcast" operations of (Pan & Kondor, 2022).

6 CONCLUSION AND OPEN QUESTIONS

Theorem 5.4 provides a basis for permutation equivariant linear layers designed to be computationally efficient, since the tensors making up the basis are constructed as Kronecker products. One practical avenue for future work would be using the theorem to implement permutation equivariant linear layers for any user-specified m, m' (to the best of our knowledge, the implementations of our references hard-code paramterizations of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ for specific values of m, m'. Another direction would be to use a presentation of the partition algebra using a subset of the diagram basis as generators (see (Jones, 1994, §3), (Halverson & Ram, 2005, Thm 1.11)) to obtain a relatively small but still expressive subspace of the permutation equivariant linear layers, as suggested in (Kim et al., 2021, §6).

[1] Explicitly, if $S_i = \{s_{i1}, \ldots, s_{i|S_i|}\}$ then in Eq. (5.1) we can use the permutation sending $(1, \ldots, l) \mapsto (s_{11}, \ldots, s_{1|S_1|}, \ldots, s_{n1}, \ldots, s_{n|S_n|})$

REFERENCES

Georgia Benkart and Tom Halverson. Partition Algebras and the Invariant Theory of the Symmetric Group, September 2017.

A Bogatskiy, Sanmay Ganguly, Thomas Kipf, Risi Kondor, David W. Miller, Daniel Murnane, Jan T. Offermann, Mariel Pettee, Phiala E. Shanahan, Chase Owen Shimmin, and Savannah Thais. Symmetry group equivariant architectures for physics. *ArXiv*, abs/2203.06153, 2022.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, May 2021.

Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, April 1980. ISSN 1432-0770. doi: 10.1007/BF00344251.

Tom Halverson and Arun Ram. Partition algebras. *European Journal of Combinatorics*, 26(6): 869–921, August 2005. ISSN 0195-6698. doi: 10.1016/j.ejc.2004.06.005.

Jason Hartford, Devon Graham, Kevin Leyton-Brown, and Siamak Ravanbakhsh. Deep Models of Interactions Across Sets. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 1909–1918. PMLR, July 2018.

V.F.R. Jones. The Potts model and the symmetric group. *Subfactors: Proceedings of the Taniguchi Symposium on Operator Algebras (Kyuzeso, 1993)*, 1994.

Jinwoo Kim, Saeyoon Oh, and Seunghoon Hong. Transformers generalize deepsets and can be extended to graphs and hypergraphs. In *Neural Information Processing Systems*, 2021.

Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. doi: 10.1162/neco.1989.1.4.541.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and Basis Invariant Networks for Spectral Graph Representation Learning, September 2022.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

P.P. Martin. *Potts Models and Related Problems in Statistical Mechanics*. Series on Advances in Statistical Mechanics. World Scientific, 1991. ISBN 978-981-02-0075-6.

P. P. Nikitin, N. V. Tsilevich, and A. M. Vershik. On the Decomposition of Tensor Representations of Symmetric Groups. *Algebras and Representation Theory*, 22(4):895–908, August 2019. ISSN 1572-9079. doi: 10.1007/s10468-018-9804-6.

Horace Pan and Risi Kondor. Fourier Bases for Solving Permutation Puzzles. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pp. 172–180. PMLR, March 2021.

Horace Pan and Risi Kondor. Permutation Equivariant Layers for Higher Order Interactions. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 5987–6001. PMLR, May 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Edward Pearce-Crump. Brauer's Group Equivariant Neural Networks, December 2022.

Edward Pearce-Crump. Connecting Permutation Equivariant Neural Networks and Partition Diagrams, January 2023.

J.P. Serre. *Linear Representations of Finite Groups*. Graduate Texts in Mathematics. Springer New York, 1977. ISBN 978-1-4684-9458-7.

Teague Sterling and John J. Irwin. Zinc 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015. doi: 10.1021/acs.jcim.5b00559. URL https://doi.org/10.1021/acs.jcim.5b00559. PMID: 26479676.

Erik Henning Thiede, Truong Son Hy, and Risi Kondor. The general theory of permutation equivarant neural networks and higher order graph variational encoders, April 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, December 2017.

Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Neural Information Processing Systems*, 2021.

Edward Wagstaff, Fabian B. Fuchs, Martin Engelcke, Michael A. Osborne, and Ingmar Posner. Universal approximation of functions on sets. *ArXiv*, abs/2107.01959, 2021.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep Sets, April 2018.

A Invariants of permutation representations

In this section, we give a concise proof of Theorem 3.10 using a couple of elementary pieces of representation theory and a combinatorial lemma.

Definition A.1. Let G be a finite group and let X be a finite set with a G-action. The **associated permutation representation**, denoted $\mathbb{R}X$, is the vector space with one basis element e_x for each $x \in X$ and the G-action defined by $g \cdot e_x = e_{gx}$.

The Σ_n representation $\mathbb{R}^n$ is a permutation representation: $\mathbb{R}^n = \mathbb{R}\{1, \dots, n\}$. In general, if X_1 and X_2 are finite sets with actions of a finite group G, then there is an isomorphism of G-representations $\mathbb{R}(X_1 \times X_2) = \mathbb{R}X_1 \otimes \mathbb{R}X_2$ (for essentially the same reason that the basic tensors $e_i \otimes e_j$ for $i = 1, \dots, m$, $j = 1, \dots n$ form a basis for $\mathbb{R}^m \otimes \mathbb{R}^n$). It follows that $(\mathbb{R}^n)^{\otimes l} = \mathbb{R}(\{1, \dots, n\}^l)$ is the permutation representation associated to the set $\{1, \dots, n\}^l$ with the "diagonal" Σ_n action.

Lemma A.2 (cf. (Jones, 1994, §1), (Benkart & Halverson, 2017, §5.2))**.** *Let G be a finite group and let X be a finite set with a G-action. Then there is a natural vector space isomorphism $\mathbb{R}(X/G) \xrightarrow{\simeq} (\mathbb{R}X)^G$, where X/G are the orbits of X, defined as follows: for each orbit $Gx \in X/G$, send the basis vector e_{Gx} to $\sum_{x' \in Gx} e_{x'}$.*

In particular, to compute $((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$, and thus by Corollary 3.6 the spaces of permutation equivariant linear layers $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ for all $m + m' = l$, all we need to do is compute the orbits of the diagonal Σ_n action on $\{1, \dots, n\}^l$ — and that is precisely the content of Lemma 3.8.

B Proofs

Proof of Lemma 3.5. In general, for any two real vector spaces V, W there is an isomorphism $V^\vee \otimes W \simeq \mathrm{Hom}(V, W)$ sending a basic tensor $\lambda \otimes w$ to the linear map $\varphi : V \to W$ defined by $\varphi(v) = \lambda(v) \cdot w$. In our case, this shows that

$$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'}) \simeq \left((\mathbb{R}^n)^{\otimes m}\right)^\vee \otimes (\mathbb{R}^n)^{\otimes m'}. \tag{B.1}$$

As $(\mathbb{R}^n)^{\otimes m}$ is a real representation of the finite group Σ_n it admits an equivariant Euclidean inner product: the natural, explicit one to use is simply defined on standard basis tensors as

$$\langle e_{i_1} \otimes \cdots \otimes e_{i_m}, e_{j_1} \otimes \cdots \otimes e_{j_m} \rangle = \prod_{k=1}^{m} \delta_{i_k j_k}. \tag{B.2}$$

Such an inner product is equivalent to an isomorphism $(\mathbb{R}^n)^{\otimes m} \simeq \left((\mathbb{R}^n)^{\otimes m}\right)^\vee$ by the map $v \mapsto \langle v, - \rangle$. Hence

$$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'}) \simeq (\mathbb{R}^n)^{\otimes m} \otimes (\mathbb{R}^n)^{\otimes m'} = (\mathbb{R}^n)^{\otimes (m+m')} \tag{B.3}$$

and taking invariants on both sides completes the proof. □

Proof of Corollary 3.6. Lemma 3.5 gives us an invertible map from $\left((\mathbb{R}^n)^{\otimes (m+m')}\right)^{\Sigma_n}$ to $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$. Then the parameterization basis of $(\mathbb{R}^n)^{\otimes l}$ is mapped to a parametrization basis of $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$. □

Proof of Lemma A.2. We note that the lemma is essentially (Serre, 1977, Ex. 2.6 (a)). A $\sum_{x \in X} c_x x \in \mathbb{R}X$ lies in the invariant subspace $(\mathbb{R}X)^G$ if and only if for each $g \in G$

$$\sum_{x \in X} c_x x = g \cdot \sum_{x \in X} c_x x = \sum_{x \in X} c_x gx = \sum_{x \in X} c_{g^{-1}x} x \tag{B.4}$$

where in the last step we have reindexed the sum. The condition that $c_x = c_{g^{-1}x}$ for all $g \in G, x \in X$ says precisely that the coefficients c_x are constant on G-orbits. In other words, if c_{Gx} is the common value of the $c_{x'}$ as x' runs over the orbit Gx then

$$\sum_{x \in X} c_x x = \sum_{Gx \in X/G} c_{Gx} \sum_{x' \in Gx} e_{x'} \tag{B.5}$$

showing that the map defined in Lemma A.2 is surjective. It is clearly injective, since for example the orbit sums $\sum_{x' \in Gx} e_{x'}$ are pairwise orthogonal. □

Proof of Lemma 3.8. Since the proof of Lemma 3.8 for any l is identical to its proof in the l-even case considered by prior work on partition algebras, we omit a proof and instead provide references where the l-even case is discussed: see (Jones, 1994, §1), (Halverson & Ram, 2005, §3) and (Benkart & Halverson, 2017, §5.1). □

Proof of Lemma 4.1. We note that in the essence of this proof is the same as the proof that a product of symmetric polynomials is symmetric.

Tensor (i.e. Kronecker) product defines a bilinear map

$$(\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'} \to (\mathbb{R}^n)^{\otimes l} \tag{B.6}$$

sending (v, w) to $v \otimes w$. Taking Σ_n-invariants gives a map

$$\left((\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \to \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{B.7}$$

Finally there is an inclusion $\left((\mathbb{R}^n)^{\otimes p}\right)^{\Sigma_n} \times \left((\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \subseteq \left((\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n}$ since the left hand side consists of pairs of tensors (v, w) invariant to the action of independent permutations $\sigma, \tau \in \Sigma_n$ as $(\sigma \cdot v, \tau \cdot w)$, and this condition is stronger than invariance to the diagonal action of a single permutation σ as $(\sigma \cdot v, \sigma \cdot w)$. □

Lemma B.8. *Let V be a real vector space and let*

$$\{v_\alpha \in V \mid \alpha \in \mathcal{A}\} \tag{B.9}$$

be a set of vectors in V indexed by a partially ordered set $\mathcal{A}$. Suppose that for every $\alpha \in A$ there exists a linear functional $\lambda_\alpha : V \to \mathbb{R}$ with the property that

$$\lambda_\alpha(v_\alpha) \neq 0 \text{ and } \lambda_\alpha(v_\beta) = 0 \text{ unless } \beta \preceq \alpha, \tag{B.10}$$

where $\preceq$ denotes the partial order on $\mathcal{A}$. Then, $\{v_\alpha \in V \mid \alpha \in \mathcal{A}\}$ is linearly independent.

Proof. Consider an equation of the form

$$\sum_{\alpha\in\mathcal{A}} c_\alpha v_\alpha = 0 \tag{B.11}$$

where all but finitely many of the c_α are 0. Suppose towards contradiction that some $c_\alpha \neq 0$, and let

$$\mathcal{B} = \{\alpha \in \mathcal{A} \,|\, c_\alpha \neq 0\} \subseteq \mathcal{A}. \tag{B.12}$$

By hypothesis, $\mathcal{B}$ is a nonempty finite partially ordered set, and as such it has at least one minimal element, i.e. an $\alpha^* \in \mathcal{B}$ such that if $\alpha \in \mathcal{B}$ and $\alpha \preceq \alpha^*$ then $\alpha = \alpha^*$. It follows that $\lambda_{\alpha^*}(v_\alpha) = 0$ for $\alpha \in \mathcal{B} \setminus \{\alpha^*\}$, hence applying λ_{α^*} to Eq. (B.11) results in

$$c_{\alpha^*}\lambda_{\alpha^*}(v_{\alpha^*}) = 0 \tag{B.13}$$

and thus $c_{\alpha^*} = 0$ (since $\lambda_{\alpha^*}(v_{\alpha^*}) \neq 0$), a contradiction. ☐

Lemma B.14. *For any set partition $\mathcal{P}$ of $\{1,\ldots,l\}$ the vector $d_\mathcal{P}$ is the sum of the $e_J = e_{j_1}\otimes\cdots\otimes e_{j_l}$ over all indices $J = (j_1,\ldots,j_l)$ that are constant on every set occurring in the partition $\mathcal{P}$ — equivalently, $d_\mathcal{P} = \sum_{\mathcal{P}\preceq\mathcal{Q}} e_\mathcal{Q}$ where $\mathcal{Q} \preceq \mathcal{P}$ if and only if $\mathcal{Q}$ refines the partition $\mathcal{P}$.*

Proof. After permuting the l tensor factors, we may reduce to the case where $\mathcal{P}$ has the form

$$\mathcal{P} = \coprod_{i=1}^{n}\{\sum_{j<i} p_j + 1,\ldots,\sum_{j<i} p_j + p_i\} \tag{B.15}$$

where $p_1,\ldots,p_n \in \mathbb{N}$ and $\sum_{i=1}^n p_i = l$. Then by direct calculation

$$d_\mathcal{P} = \sum_{j_1,\ldots,j_n=1}^{n} e_{\underbrace{j_1,\ldots,j_1}_{p_1\text{times}}} \otimes e_{\underbrace{j_2,\ldots,j_2}_{p_2\text{times}}} \otimes\cdots\otimes e_{\underbrace{j_n,\ldots,j_n}_{p_n\text{times}}}, \tag{B.16}$$

and evidently the indices $(\underbrace{j_1,\ldots,j_1}_{p_1\text{times}},\ldots,\underbrace{j_n,\ldots,j_n}_{p_n\text{times}})$ occurring in the sum are exactly those constant on each of the sets $\{\sum_{j<i} p_j + 1,\ldots,\sum_{j<i} p_j + p_i\}$. ☐

Proof of Theorem 5.4. Since both the $e_\mathcal{P}$ and the $d_\mathcal{P}$ are indexed by set partitions of $\{1,\ldots,l\}$ into at most n non-empty subsets they have the same cardinality, and we already know the $e_\mathcal{P}$ are a basis for $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$. Thus, it will suffice to show the $d_\mathcal{P}$ are linearly independent. We will prove this by exhibiting a set of linear functionals

$$\{\lambda_\mathcal{P} : \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \to \mathbb{R} \,|\, \mathcal{P} \text{ is a set partition of } \{1,\ldots,l\}\} \tag{B.17}$$

such that

$$\lambda_\mathcal{P}(d_\mathcal{P}) \neq 0 \text{ and } \lambda_\mathcal{P}(d_\mathcal{Q}) = 0 \text{ unless } \mathcal{Q} \preceq \mathcal{P}, \tag{B.18}$$

where $\mathcal{Q} \preceq \mathcal{P}$ if and only if $\mathcal{Q}$ refines the partition $\mathcal{P}$, satisfying the properties of Lemma B.8 below. Explicitly, for each partition $\mathcal{P}$ choose an index $I_\mathcal{P} = (i_1,\ldots,i_l) \in \{1,\ldots,n\}^l$ such that $\mathcal{P}$ is the partition associated to $I_\mathcal{P}$ as in Lemma 3.8, and let $e_{I_\mathcal{P}} = e_{i_1}\otimes\cdots\otimes e_{i_l} \in (\mathbb{R}^n)^{\otimes l}$. Equivalently, $e_{I_\mathcal{P}}$ is one of the standard basis vectors for $(\mathbb{R}^n)^{\otimes l}$ occurring in the orbit sum defining $e_\mathcal{P}$. Then define $\lambda_\mathcal{P}$ as dot product with $e_{I_\mathcal{P}}$:

$$\lambda_\mathcal{P}(v) = \langle e_{I_\mathcal{P}}, v\rangle \text{ for } v \in \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{B.19}$$

First, since by definition $I_\mathcal{P}$ is constant on each set of the partition $\mathcal{P}$, our characterization of $d_\mathcal{P}$ in Lemma B.14 gives $\langle e_{I_\mathcal{P}}, d_\mathcal{P}\rangle = 1$. Next, if $\mathcal{Q} \not\preceq \mathcal{P}$, then since $\mathcal{Q}$ doesn't refine $\mathcal{P}$ writing

$$\mathcal{P} = \{S_1,\ldots,S_n\} \text{ and } \mathcal{Q} = \{T_1,\ldots,T_n\} \tag{B.20}$$

there must be a non-empty T_i such that $T_i \not\subseteq S_j$ for all j. Since $T_i \subseteq \bigcup_j S_j$, there must be distinct $S_j, S_k \in \mathcal{P}$ with $S_j \cap T_i \neq \emptyset, S_k \cap T_i \neq \emptyset$. By design the index $I_\mathcal{P}$ takes distinct values on S_j and S_k, but for every $e_J = e_{j_1}\otimes\cdots\otimes e_{j_l}$ occurring in $d_\mathcal{Q}$ with non-zero coefficient the index J is constant on T_i. Thus $\langle e_{I_\mathcal{P}}, d_\mathcal{Q}\rangle = 0$. ☐

Proof of Proposition 5.5. This follows from Lemma B.14 and (Jones, 1994, p. 263) (see also (Benkart & Halverson, 2017, §4.2-3)). ☐

C AN ALGORITHM FOR MULTIPLICATION WITH THE TENSORS $d_{\mathcal{P}}$

In this section we describe how to apply one of the tensors $d_{\mathcal{P}} \in \big((\mathbb{R}^n)^{\otimes(m+m')}\big)^{\Sigma_n} \simeq \mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ appearing in Theorem 5.4 to a tensor $v \in (\mathbb{R}^n)^{\otimes m}$. We will work in the usual bases for these tensor products, and use notation of the form

$$v = \sum_{i_1,\ldots,i_m=1}^{n} v_{i_1\ldots i_m} e_{i_1} \otimes \cdots \otimes e_{i_m}. \tag{C.1}$$

Given another tensor

$$w = \sum_{j_1,\ldots,j_l=1}^{n} w_{j_1\ldots j_l} e_{j_1} \otimes \cdots \otimes e_{j_l} \in (\mathbb{R}^n)^{\otimes l} \tag{C.2}$$

and ordered tuples of indices T and S with $\mathrm{set}(T) \subseteq \{1,\ldots,m\}, \mathrm{set}(S) \subseteq \{1,\ldots,l\}$ and tuple length $d \le \min\{m,l\}$, we will use the notation $\mathrm{dot}(v,w,(S,T))$ to denote the tensor contraction operation implemented in PyTorch as $\texttt{tensordot}(v,w,\texttt{dims} = (S,T))$ (Paszke et al., 2019). The special case where $T = (m, m-1, \ldots, m-d+1), S = (1,\ldots,d)$ will be abbreviated as $\mathrm{dot}(v,w,d)$. We refer to the documentation at https://pytorch.org/docs/stable/index.html for further details.

Unravelling Lemma 3.5, we see that our goal is to calculate $\mathrm{dot}(v, d_{\mathcal{P}}, m)$. The point we wish to make is that for many partitions $\mathcal{P}$ this can be accomplished with multiple contractions over fewer than m indices, and moreover the contractions occurring can be replaced with sums and indexing operations.

The contraction $\mathrm{dot}(v, d_{\mathcal{P}}, m)$ is invariant to permutations of the tensor factors of v and the first m tensor factors of $d_{\mathcal{P}}$. Moreover, we are free to permute the last m' factors of $d_{\mathcal{P}}$ provided we apply the inverse permutation to the m' factors of $\mathrm{dot}(v, d_{\mathcal{P}}, m)$. These two observations allow us to reduce to the case where $\mathcal{P}$ has the form

$$(\coprod_{i=1}^{a} S_i) \coprod (\coprod_{i=1}^{b} T_i) \coprod (\coprod_{i=1}^{c} B_i) \tag{C.3}$$

where

- $a + b + c = n$,
- the S_i are consecutive and contiguous sets of indices in $\{1,\ldots,m\}$ beginning at 1 and ending at say p,
- the B_i are consecutive and contiguous sets of indices in $\{m+1,\ldots,m+m'\}$ beginning at say $p+p'$ and ending at $m+m'$, and
- the T_i partition $\{p, p+1, \ldots, p+p'\}$, and moreover they decompose as $T_i = T_i' + T_i''$ where $T_1', \ldots, T_b'$ are consecutive and contiguous sets of indices in $\{p+1,\ldots,m\}$ and $T_b'', \ldots, T_1''$ are consecutive and contiguous sets of indices in $\{p+p'+1,\ldots,m+m'\}$.

It follows that

$$\begin{aligned} d_{\mathcal{P}} &= e_{\{\{1,\ldots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|S_a|\}\}} \\ &\quad \otimes \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}}) \\ &\quad \otimes e_{\{\{1,\ldots,|B_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|B_c|\}\}} \end{aligned} \tag{C.4}$$

where Ψ permutes tensor factors according to a certain permutation of $\{p, p+1, \ldots, p+p'\}$ (the one separating each T_i into the subsets T_i', T_i'').

We now make repeated use of two simple calculations: first,

$$\mathrm{dot}(v,w,d) = \mathrm{dot}(v,w',d) \otimes w'' \text{whenever } w = w' \otimes w'' \text{ where } w' \text{ has} \ge d \text{ indices.} \tag{C.5}$$

That is, w'' can be extracted from the dot. On the other hand,

$$\mathrm{dot}(v,w,d) = \mathrm{dot}(\mathrm{dot}(v,w',d'),w'',d-d') \text{whenever } w = w' \otimes w'' \text{ where } w' \text{ has } d' \le d \text{ indices} \tag{C.6}$$

In particular, Eq. (C.5) applies to the factors e_{B_i} in Eq. (C.4), giving

$$\begin{aligned}&\mathrm{dot}(v, d_{\mathcal{P}}, m)\\&= \mathrm{dot}(v, e_{\{\{1,\ldots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|S_a|\}\}} \otimes \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}}), m)\\&\otimes e_{\{\{1,\ldots,|B_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|B_c|\}\}}.\end{aligned} \tag{C.7}$$

An explicit calculation shows that the tensor operation $x \mapsto x \otimes e_{\{1,\ldots,p\}}$ simply creates a tensor with p more indices than x and places copies of x along indices of the form $\ldots, \underbrace{i, i, \ldots, i}_{p \text{ times}}$, with zeros elsewhere. Thus once

$$\mathrm{dot}(v, e_{\{\{1,\ldots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|S_a|\}\}} \otimes \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}}), m)$$

is computed, multiplication with the $e_{\{\{1,\ldots,|B_i|\}\}}$ can be accomplished efficiently with indexing operations.[2]

Equation (C.6) shows that contraction with the $e_{\{\{1,\ldots,|S_i|\}\}}$ can be carried out "one i at a time," from left to right. Based on our conventions for dot, a contraction $\mathrm{dot}(v, e_{\{\{1,\ldots,p\}\}}, p)$ yields a tensor with $m-p$ indices, with $(i_1, \ldots, i_{m-p})$-th entry

$$\sum_{j=1}^{n} v_{i_1 \ldots i_{m-p} \underbrace{j \ldots j}_{p \text{ times}}} \tag{C.8}$$

Hence these contractions can be implemented with index operations and summation.[3]

Finally, the tensor $\Psi(e_{T_1} \otimes \cdots \otimes e_{T_b})$ must be dealt with. We claim that dot with this tensor is a *transfer operation* as described in (Pan & Kondor, 2022, §5). Indeed, they define transfer operations as those corresponding to sets in the partition $\mathcal{P}$ having non-empty intersection with both $\{1, \ldots, m\}$ and $\{m+1, \ldots, m+m'\}$, which is exactly the role played by the T_i. Since we have already addressed how to multiply with the $e_{\{\{1,\ldots,|S_i|\}\}}$ and $e_{\{\{1,\ldots,|B_i|\}\}}$ we may as well assume for simplicity that $d_{\mathcal{P}} = \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}})$. Then,

$$\mathrm{dot}(v, \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}}))_{\underbrace{i_1 \ldots i_1}_{|T_1''| \text{ times}} \ldots \underbrace{i_b \ldots i_b}_{|T_b''| \text{ times}}} = v_{\underbrace{i_1 \ldots i_1}_{|T_1'| \text{ times}} \ldots \underbrace{i_b \ldots i_b}_{|T_b'| \text{ times}}}. \tag{C.9}$$

Clearly, this is essentially an indexing operation.[4]

D Examples

We look at the case where $m = m' = 1$ and $n \geq 2$. Here the space of permutation equivariant linear layers is $\left((\mathbb{R}^n)^{\otimes 2}\right)^{\Sigma_n}$, i.e. matrices invariant under simultaneous permutation of rows and columns, and these corresponding to equivariant maps $\mathbb{R}^n \to \mathbb{R}^n$ by matrix-vector multiplication.

D.1 Partitions associated with index tuples

The simplest case of Lemma 3.8 occurs when $l = 2$. Here, given $(i_1, i_2) \in \{1, \ldots, n\}^2$ the associated set partition is

$$\Pi(i_1, i_2) = \begin{cases} \{\{1,2\}\} & \text{if } i_1 = i_2 \\ \{\{1\},\{2\}\} & \text{if } i_1 \neq i_2 \end{cases} \tag{D.1}$$

Let $\mathcal{P}_1 = \{\{1,2\}\}$ and $\mathcal{P}_2 = \{\{1\},\{2\}\}$. Then viewing tensors in $\mathbb{R}^n \otimes \mathbb{R}^n$ as $n \times n$ matrices,

$$\begin{aligned} e_{\mathcal{P}_1} &= \sum_i e_i \otimes e_i = I_n \text{ and} \\ e_{\mathcal{P}_2} &= \sum_{i \neq j} e_i \otimes e_j = \mathbf{1}\mathbf{1}^T - I \end{aligned} \tag{D.2}$$

[2] Observe that in contrast Kronecker multiplication with a general tensor w'' with p entries would require n^p scalar multiplications per entry of x, so e.g. if x has q indices n^{p+q} scalar multiplications in total.

[3] Observe that Eq. (C.6) reduces the cost of tensor contraction from roughly n^d to $n^{d'} + n^{d-d'}$ multiplications and additions. In the special case where $w' = e_{\{1,\ldots,d'\}}$ the term $n^{d'}$ effectively drops to n.

[4] In other words, rather than performing $n^{\sum |T_i'|}$ multiplications and additions, we are just copying arrays.

where I is the $n \times n$ identity matrix and $\mathbf{1} = [1, \ldots, 1]^T$ the 1s vector.

D.2 RANKS OF BASIS VECTORS

Observe that $\mathbf{1}\mathbf{1}^T - I$ is *not* a Kronecker product uv^T. Indeed, it has rank n, since $det(-I + \mathbf{1}\mathbf{1}^T) = (1 + \mathbf{1}^T(-I)\mathbf{1})det(I) = (1-n)n \neq 0$, whereas any uv^T has rank 1.

On the other hand, the basis of Theorem 5.4 is

$$d_{\mathcal{P}_1} = I \text{ and } d_{\mathcal{P}_2} = \mathbf{1}\mathbf{1}^T \tag{D.3}$$

and $\mathbf{1}\mathbf{1}^T$ is of course rank 1. This is of course the basis used in DeepSets (Zaheer et al., 2018) (I is the identity map, $\mathbf{1}\mathbf{1}^T$ is the vector sum multiplied by $\mathbf{1}$).