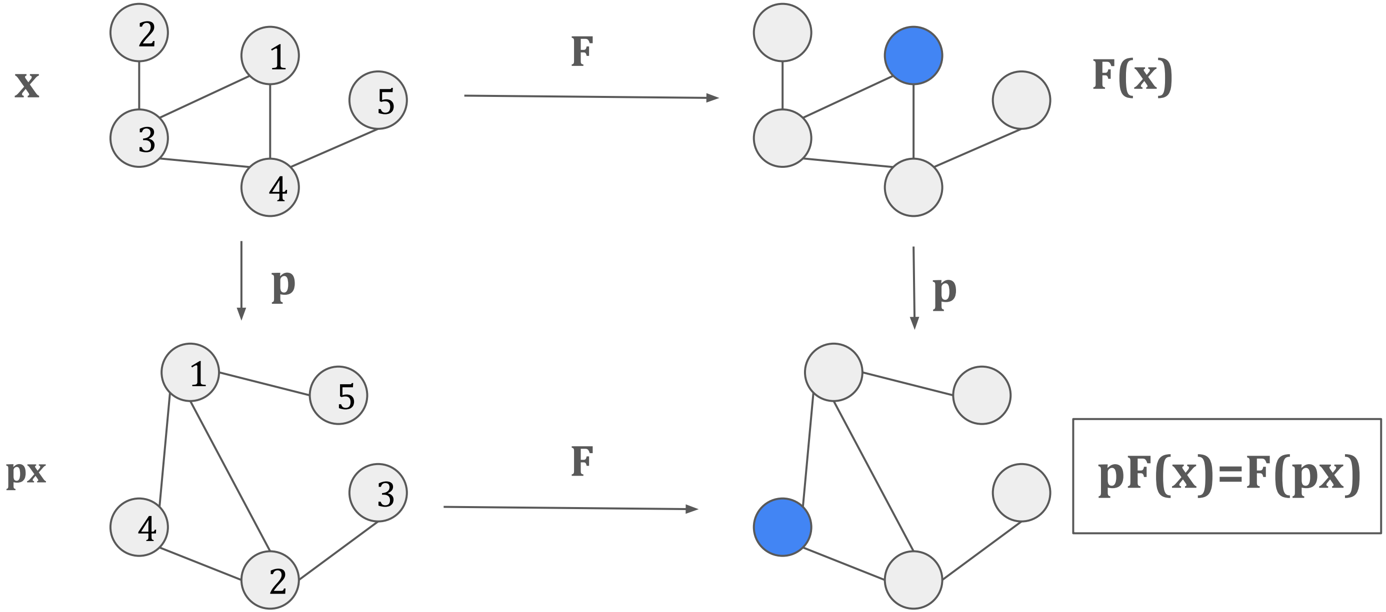
ICLR

Revisiting Multi-Permutation Equivariance Through the Lens of Irreducible Representations

Yonatan Sverdlov*, Ido Springer*, Nadav Dym

TECHNION
Israel Institute of Technology

Problem Setting: Finding Permutation Equivariant linear Layers



TECHNION
Israel Institute
of Technology

Problem Setting: Finding Permutation Equivariant linear Layers

- Find all linear layers F s.t. $F(px) = pF(x)$ for any permutation $p \in G$

TECHNION
Israel Institute
of Technology

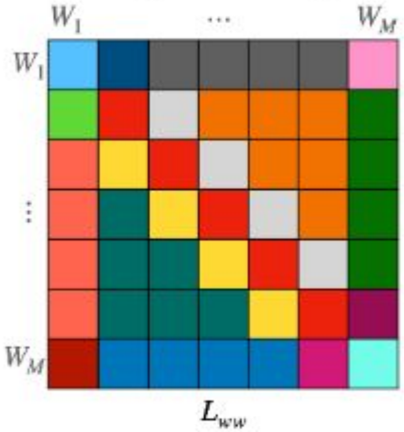
Problem Setting: Finding Permutation Equivariant Layers

- Find linear layers F s.t. $F(px) = pF(x)$ for any permutation $p \in G$
- A crucial architecture component for geometric domains
  - such as sets, graphs and weight spaces.

TECHNION
Israel Institute
of Technology

Problem Setting: Finding Permutation Equivariant Layers

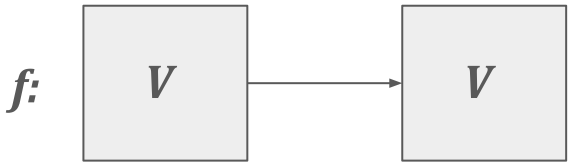
- Find linear layers F s.t. $F(px) = pF(x)$ for any permutation $p \in G$
- Traditionally this was done by finding explicit parameter sharing.
- However explicit parameter sharing tend to be tedious.



TECHNION
Israel Institute of Technology

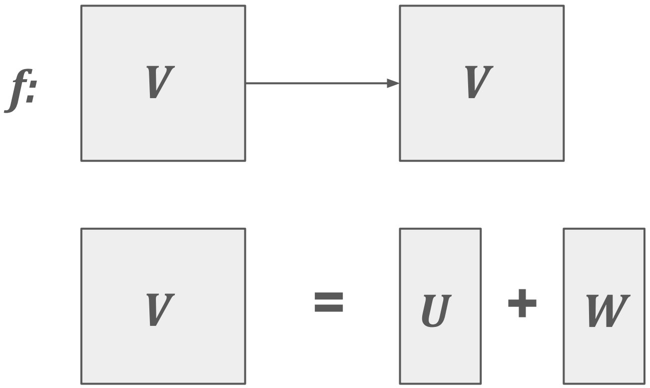
Illustration: Navon et al., Equivariant Architectures for Learning in Deep Weight Spaces

Irreducible Representations and Schur's Lemma



G-equivariant layer

Irreducible Representations and Schur’s Lemma

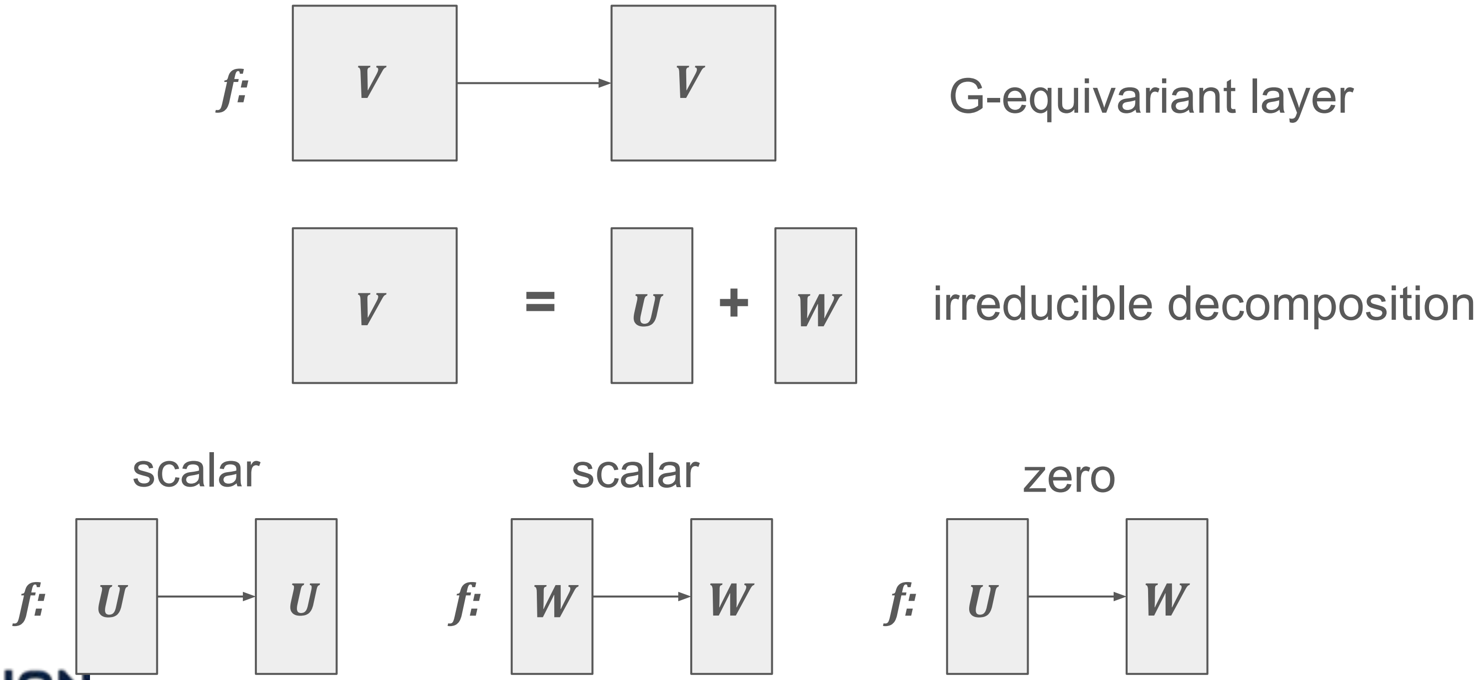


G-equivariant layer

irreducible decomposition

TECHNION
Israel Institute
of Technology

Irreducible Representations and Schur's Lemma



Re-deriving Permutation Equivariant Layers

- We re-derive known equivariant characterizations using irreducible decompositions and Schur's lemma
- Case studies:
  - DeepSets (Zaheer et al.)
  - Graph Invariant Networks (Maron et al.)
  - Deep Weight Spaces Networks (Navon et al.)
  - Wreath-equivariant layers

TECHNION
Israel Institute of Technology

Case Study: DeepSets

- DeepSets (Zaheer et al. 2017): S_n acting on R^n.
- two irreps: scalars and sum to zero
- decompose x in R^n to x=x_bar*1n + (x-x_bar*1n)
- Now we get Tx = a*x_bar*1n + b*(x-x_bar*1n) by Schur
- this is exactly the original result of DeepSets!
- We do the same for IGN and DWS

TECHNION
Israel Institute
of Technology

Wreath-Equivariant Layers

- Sets of *unaligned* symmetric elements $(\mathbf{v}_1,\ldots,\mathbf{v}_k)$
- Equivariance with respect to the joint action of:
  - k-tuple of group elements $(\mathbf{g}_1,\ldots,\mathbf{g}_k)$ on each coordinate independently
  - a permutation $\boldsymbol{\tau} \in \boldsymbol{S}_k$ of the k-tuple
- This corresponds to the action of the wreath product $\mathbf{G} \wr \boldsymbol{S}_k$ on $\mathbf{V}^k$
- We seek for wreath equivariant linear layers
- Applications: alignment problems, hierarchical structures

TECHNION
Israel Institute
of Technology

Full Characterization of Wreath-Equivariant Layers

Theorem 5.2. *Let $\mathcal{V}$ be a real representation of a finite group $\mathcal{G}$. and let $e_1, \ldots, e_s$ be a basis to the subspace $\mathcal{V}_{fixed}$. Let $\langle \cdot, \cdot \rangle$ be a $\mathcal{G}$ invariant inner product on $\mathcal{V}$. Then every linear equivariant map $L : \mathcal{V}^k \to \mathcal{V}^k$ is of the form*

$$L(v_1, \ldots, v_k) = \sum_{i,j=1}^{s} a_{ij} \left(\sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j, \ldots, \sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j \right) + \left(\hat{L}(v_1), \ldots, \hat{L}(v_k) \right) \qquad (9)$$

where $\hat{L} : \mathcal{V} \to \mathcal{V}$ is a linear equivariant map, and a_{ij} are real numbers. Conversely, every linear mapping of the form defined in equation 9 is equivariant.

TECHNION
Israel Institute of Technology

Full Characterization of Wreath-Equivariant Layers

Theorem 5.2. *Let $\mathcal{V}$ be a real representation of a finite group $\mathcal{G}$. and let $e_1, \ldots, e_s$ be a basis to the subspace $\mathcal{V}_{fixed}$. Let $\langle \cdot, \cdot \rangle$ be a $\mathcal{G}$ invariant inner product on $\mathcal{V}$. Then every linear equivariant map $L : \mathcal{V}^k \to \mathcal{V}^k$ is of the form*

$$L(v_1, \ldots, v_k) = \sum_{i,j=1}^{s} a_{ij} \left(\sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j, \ldots, \sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j \right) + \left(\hat{L}(v_1), \ldots, \hat{L}(v_k) \right) \tag{9}$$

where $\hat{L} : \mathcal{V} \to \mathcal{V}$ is a linear equivariant map, and a_{ij} are real numbers. Conversely, every linear mapping of the form defined in equation 9 is equivariant.

siamese layers

TECHNION
Israel Institute of Technology

Full Characterization of Wreath-Equivariant Layers

Theorem 5.2. *Let $\mathcal{V}$ be a real representation of a finite group $\mathcal{G}$. and let $e_1, \ldots, e_s$ be a basis to the subspace $\mathcal{V}_{fixed}$. Let $\langle \cdot, \cdot \rangle$ be a $\mathcal{G}$ invariant inner product on $\mathcal{V}$. Then every linear equivariant map $L : \mathcal{V}^k \to \mathcal{V}^k$ is of the form*

$$L(v_1, \ldots, v_k) = \sum_{i,j=1}^{s} a_{ij} \left(\sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j, \ldots, \sum_{\ell=1}^{k} \langle v_\ell, e_i \rangle e_j \right) + \left(\hat{L}(v_1), \ldots, \hat{L}(v_k) \right) \qquad (9)$$

where $\hat{L} : \mathcal{V} \to \mathcal{V}$ is a linear equivariant map, and a_{ij} are real numbers. Conversely, every linear mapping of the form defined in equation 9 is equivariant.

additional $\mathbf{s^2}$
non-siamese layers!

Siamese layers

s = num. of trivial irreps

TECHNION
Israel Institute
of Technology

Experiments: Graph Anomaly Detection

Find anomalous graph within $\mathbf{k}$ graphs with $\mathbf{n}$ nodes each.

Model/Noise	$\eta = 0.0$	$\eta = 0.1$
Siamese	10%	10%
DSS	97.5%	92%
SchurNet (Ours)	**100%**	**97.0%**

Table 1: Performance comparison of models at different noise levels (η).

Non-siamese layers are crucial!

TECHNION
Israel Institute
of Technology

Experiments: Wasserstein Distance Computation

Learning Wasserstein distances is a $S_n \wr S_2$ invariant task.

Dataset	Input	SchurNet	NProductNet
noisy-sphere-3	[100, 300]	**0.0389**	0.046
	[300, 500]	**0.1026**	0.158
noisy-sphere-6	[100, 300]	0.0217	**0.015**
	[300, 500]	0.0795	**0.049**
uniform	256	0.0974	**0.097**
	[200, 300]	**0.1043**	0.1089
ModelNet-small	[20, 200]	**0.0623**	0.084
	[300, 500]	**0.0738**	0.111
ModelNet-large	2048	**0.0468**	0.140
	[1800, 2000]	**0.0551**	0.162
RNAseq	[20, 200]	0.0123	**0.012**
	[300, 500]	**0.0334**	0.292

Table 2: Comparison of SchurNet and NProductNet.

TECHNION
Israel Institute of Technology

Experiments: Weight Space Alignment

- $\mathbf{v}_1, \mathbf{v}_2$ are elements in weight space
- Find the group element that optimally aligns $\mathbf{v}_1$ and $\mathbf{v}_2$.

Model	MNIST		CIFAR10	
	Acc(↓)	Loss (↓)	Acc(↓)	Loss (↓)
SchurNet (Ours)	**1.25e-5**	**0.251346**	**0.0**	**1.7822**
Siamese	1.5e-5	0.262913	1.0e-4	1.7876

Table 3: Comparison of SchurNet and Siamese models on MNIST and CIFAR10.

Non-siamese layers improve error rates.

TECHNION
Israel Institute
of Technology

ICLR

Thanks!

arXiv: https://arxiv.org/abs/2410.06665

Code: https://github.com/yonatansverdlov/SchurNet

Poster: Fri 25 Apr 3PM+08 - 5:30PM+08

read our paper

TECHNION
Israel Institute
of Technology