Neural Discovery of Permutation Subgroups

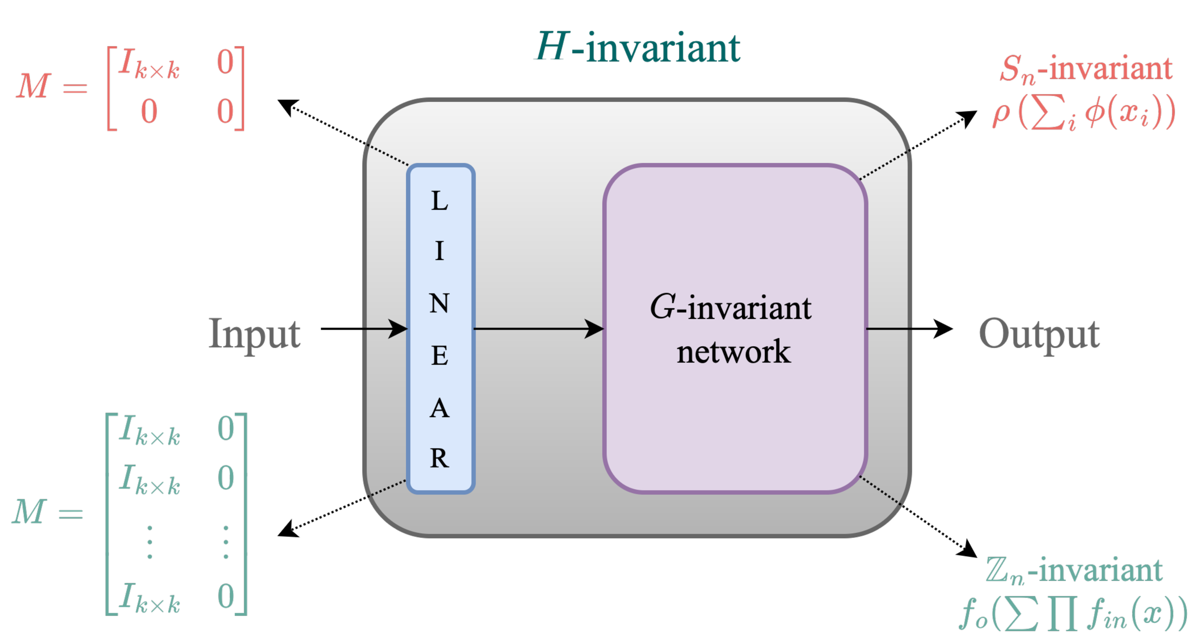
Pavan Karjol, Rohan Kashyap, Prathosh A P

Electrical and Communication Department, Indian Institute of Science

Problem statement

We consider the problem of learning an H-invariant function $f : X \to \mathbb{R}$, where $X = [0, 1]^n \subset R^n$ and H is the unknown subgroup of S_n. In general, learning such a function is intractable. However, we show that it is possible to learn such a function, i.e., discover the underlying subgroup H, where H belongs to a certain class of subgroups.

Method



We propose a general framework, i.e., G-invariant network and a linear transformation for discovering the underlying subgroup of S_n under certain conditions. Since any given G can have several such subgroups, we propose to learn the underlying subgroup H by exploiting the existing structures using a family of G-invariant functions (as mentioned in [2] for the permutation group S_n) and a learnable linear transformation.

Deep Sets

$f : X = [0, 1]^n \to \mathbb{R}$ is a permutation invariant (S_n-invariant) continuous function if it has the representation [2] given as follows:

$$f(x) = \rho\left(\sum_{i=1}^{n} \gamma(x_i)\right), \quad x = [x_1, x_2, \ldots x_n]^T$$

for some continuous outer and inner functions $\rho : \mathbb{R}^{n+1} \to \mathbb{R}$, $\gamma : [0, 1] \to \mathbb{R}^{n+1}$.

S_k - Invariance

Any S_k-invariant function ($k \le n$) ψ, can be realised using an S_n-invariant function and a linear transformation, in specific, it can be realised through:

$$\psi(x) = \left(\phi \cdot \hat{M}\right)(x) = \rho\left(\begin{bmatrix} (I - M)x \\ \sum_{i=1}^{n} \gamma\left(m_i^T x\right) \end{bmatrix}\right)$$

where $\hat{M} = \begin{bmatrix} I - M \\ M \end{bmatrix}$ and $\phi(y) = \left[y_1, \ldots, y_n, \sum_{i=1}^{n} \gamma(y_{n+i})\right]^T$

$\mathbb{Z}_k$(or D_{2k}) - Invariance

If $k|n$, any $\mathbb{Z}_k$-invariant (or D_{2k}-invariant) function ψ, can be realised using a $\mathbb{Z}_n$-invariant (or D_{2n}-invariant) function ϕ and a linear transformation, can be realised as follows:

$$\psi(x) = \left(\phi \cdot \hat{M}\right)(x), \text{ where } \hat{M} = \begin{bmatrix} M \\ I - L \end{bmatrix} \text{ for some } M, L \in \mathbb{R}^{n\times n}.$$

General Invariance

Any H-invariant function ψ can be learnt through composing a G-invariant function ϕ with a linear transformation M, i.e., $\psi = \phi \cdot M$ if the following conditions hold,

1. For any $h \in H$, $\exists g \in G$ such that $M(h \cdot x) = g \cdot (Mx)$, $\forall x \in X$
2. For any $g \in G$ such that $g \cdot (Mx) \in R(M)$, $\exists h \in H$ such that $M(h \cdot x) = g \cdot (Mx)$, $\forall x \in X$, where $R(M)$ is the range of M.

Overview

The general invariance result presents a set of conditions to be satisfied to learn any H-invariant function using a G invariant function and a linear transformation. As such, the previous results are specific cases of this result. However, they provide explicit structures of the linear transformation M. These can help design appropriate training techniques to learn the optimum M.

1. We show that we could learn any conjugate group (with respect to G) via a linear transformation and G-invariant network.
2. We extend this approach, i.e., a linear transformation and G-invariant network to different classes of subgroups such as permutation group of k (out of n) elements S_k, cyclic subgroups $\mathbb{Z}_k$ and dihedral subgroups D_{2k}.
3. We prove a general theorem that can guide us to discover other classes of subgroups.

Interpretability

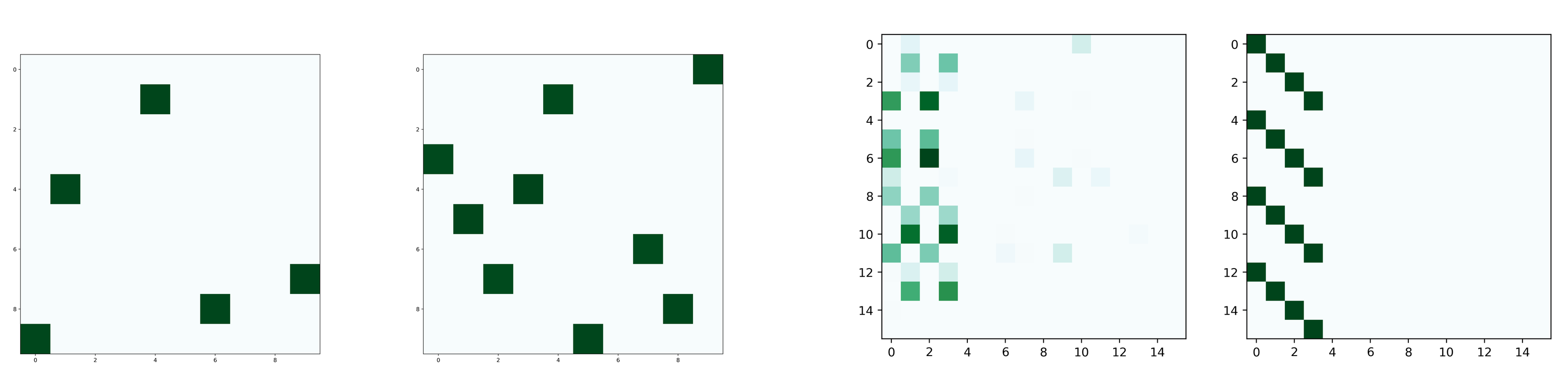


Figure 1. M Matrices for S_5 (a) and S_9 (b) after training. M matrix for $\mathbb{Z}_4 : \mathbb{Z}_{16}$(c) and Reference matrix (d).

The resulting M matrix is interpretable, and we consistently observe the expected pattern for the image-digit sum task. Note: Any row-permuted version of the matrix structure will work since the transformed space is still homeomorphic.

We observe that the M-matrix does not represent a stack of I matrices even though it nearly masks most of the irrelevant columns $(n-k)$. The former behavior (lack of exact structure) explains the difference in performance with respect to the $\mathbb{Z}_k$-invariant network, while the latter (masking behavior) describes the superior model performance compared to other baselines.

Explicit Linear Transformations

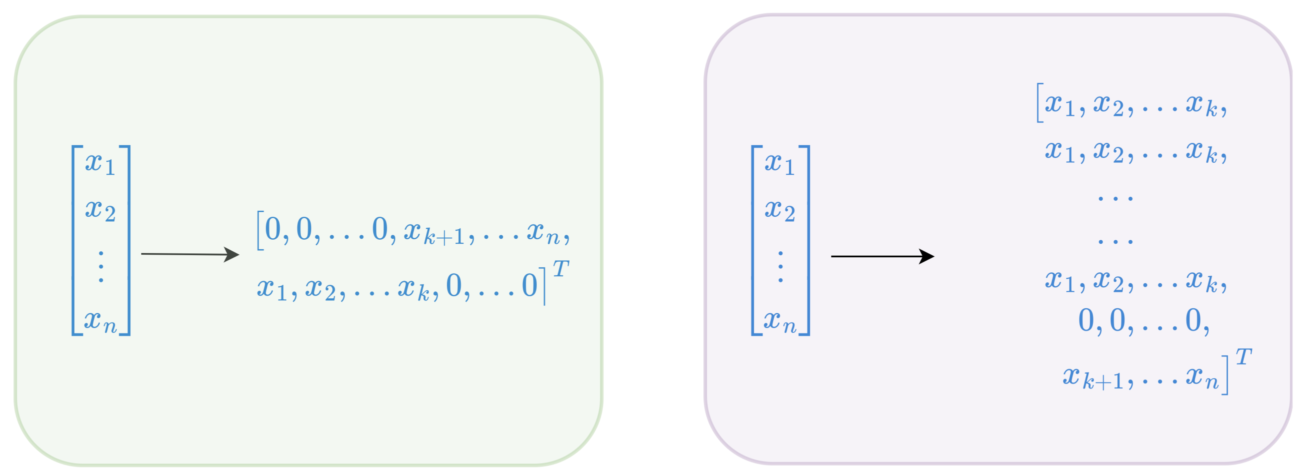


Figure 2. Linear transformations used for $\mathbb{S}_k$, $\mathbb{Z}_k$ and $\mathbb{D}_{2k}$-invariant functions.

Results

Symmetric Polynomial Regression: We evaluate the performance of our method on symmetric polynomial regression tasks as discussed in [1], primarily for subgroups of $\mathbb{Z}_{10}$ and $\mathbb{Z}_{16}$. For all our experiments, we utilize a $\mathbb{Z}_n$-invariant neural network with a Sum-Product layer as discussed in [1] and a linear layer.

Table 1. MAE $[\times 10^{-2}]$ for $\mathbb{Z}_5 : \mathbb{Z}_{10}$

Method	Train	Validation	Test
$\mathbb{Z}_5$-invariant	2.65 ± 0.91	7.32 ± 0.55	7.53 ± 0.576
Proposed	4.48 ± 1.25	24.56 ± 6.93	24.78 ± 6.45
Conv-1D	20.90 ± 4.91	32.96 ± 1.31	32.33 ± 1.18
Simple-FC	23.86 ± 3.87	33.57 ± 2.07	33.14 ± 2.11

Image-Digit Sum: This task aims to find the sum of k digits using the MNISTm dataset. Our method outperforms the each of the baseline networks for any given subgroups. We observe that the proposed method outperforms the LSTM baseline and is competitive with respect to the Deep Sets method (k input images) when the underlying subgroup S_k is known.

Table 2. MAE $[\times 10^{-2}]$ for Image Digit-Sum task

Method	S_1	S_3	S_5	S_7	S_9
Deep Sets-S_k	5.61 ± 0.35	7.66 ± 0.26	8.02 ± 0.2	7.68 ± 0.43	6.97 ± 0.39
Proposed	5.73 ± 0.39	7.78 ± 0.49	8.19 ± 0.36	7.84 ± 0.41	7.26 ± 0.58
LSTM	6.23 ± 0.53	9.65 ± 0.57	11.98 ± 0.46	13.35 ± 1.02	12.92 ± 1.42

Limitations

- Information regarding broader caterory of the subgroup is required apriori.
- Have to choose appropriate value for n.

References

[1] Piotr Kicki, Mete Ozay, and Piotr Skrzypczyński.
A computationally efficient neural network invariant to the action of symmetry subgroups.

[2] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola.
Deep sets.