The basic paradigm of deep neural networks is repeatedly composing linear layers interleaved with non-linear, element-wise activation functions to create effective predictive models. When trying to learn a function (task) $f$ that is known to be invariant to some group of symmetries $G$ (i.e., $G$-invariant function), it is common to use linear layers that respect this symmetry, namely, invariant and/or equivariant linear layers. Networks with invariant/equivariant linear layers with respect to some group $G$ will be referred here as $G$-invariant networks, e.g., Figure 1.

**Goal**

When using these networks one important question arises: Can a $G$-invariant network approximate on arbitrary continuous $G$-invariant function?

The proof of Theorem 1 is constructive: given a $G$-invariant function $f$, it builds an $f$-approximating $G$-invariant network with hidden tensors $X \in \mathbb{R}^{d_0}$ of order $d$, where $d = d(G)$ is a natural number depending on the group $G$. It can be shown that $d = \frac{n}{2}$ for all $G \leq S_n$.

The main idea is to approximate a finite generating set of the ring of $G$-invariant polynomials, which can be shown to be dense in the space of continuous $G$-invariant functions on a compact domain.

### A lower bound on equivariant layer order

We note that even using tensors $X \in \mathbb{R}^{d_0}$ with order $d = 2$ could already be computationally challenging. We therefore ask whether we can use $G$-invariant networks with lower order tensors without sacrificing approximation power.

We show this is generally not the case; for some subgroups a minimal order of $d \geq 3$ is needed. Specifically, we prove the following for $G = A_n$, the alternating group:

**Theorem**

If an $A_n$-invariant network has the universal approximation property then it consists of tensors of order at least $\frac{n}{2}$.

The basic idea behind this proof is that both $S_n$ and $A_n$ give rise to the same linear equivariant layers for low order tensors. This implies that any $A_n$-invariant network is also $S_n$-invariant. We use this fact in order to show that an $A_n$-invariant network that uses only low order tensors cannot approximate the Vandermonde polynomial $V(x) = \prod_{i=0}^{n-1} (x_i - x)$ (which is $A_n$-invariant but not $S_n$-invariant).

### University of first order networks

Although in general we cannot expect universal approximation of $G$-invariant networks with inner tensor order smaller than $\frac{n}{2}$, it is still possible that for specific groups of interest we can prove approximation power with more efficient (i.e., lower order inner tensor) $G$-invariant networks. Of specific interest are $G$-invariant networks that use only first order tensors.

**Theorem**

Let $G \leq S_n$. If first order $G$-invariant networks are universal, then $\|a^n\|_H < \frac{n}{2} |G|/|H|$ for any strict super-group $H < S_n$.

$\|a^n\|_H$ is the number of equivalence classes of $[a]^n$ defined by the relation: $(i_1, i_2) \sim (j_1, j_2)$ if $j_1 - j_2 = g(i_1) - g(i_2)$, $\ell = 1, 2$ for some $g \in G$. Intuitively, this condition asks that super-groups of $G$ have strictly better separability of the double index space $[a]^n$.

### References