Drawback of Enforcing Equivariance and its Compensation via the Lens of Expressive Power

Yuzhu Chen* *cyzkrau@mail.ustc.edu.cn*
University of Science and Technology of China

Tian Qin* *tqin@mail.ustc.edu.cn*
University of Science and Technology of China

Xinmei Tian *xinmei@ustc.edu.cn*
University of Science and Technology of China

Fengxiang He *F.He@ed.ac.uk*
University of Edinburgh

Dacheng Tao *dacheng.tao@ntu.edu.sg*
Nanyang Technological University

Reviewed on OpenReview: *https://openreview.net/forum?id=z5bJ44Brc4*

Abstract

Equivariant neural networks encode the intrinsic symmetry of data as an inductive bias, which has achieved impressive performance in wide domains. However, the understanding to their expressive power remains premature. Focusing on 2-layer ReLU networks, this paper investigates the impact of enforcing equivariance constraints on the expressive power. By examining the boundary hyperplanes and the channel vectors, we constructively demonstrate that enforcing equivariance constraints could undermine the expressive power. Naturally, this drawback can be compensated for by enlarging the model size – we further prove upper bounds on the required enlargement for compensation. Surprisingly, we show that the enlarged neural architectures have reduced hypothesis space dimensionality, implying even better generalizability.

1 Introduction

In scaling up the capacity of machine learning models, leveraging the intrinsic symmetry in data, if it exists, would be a much more efficient approach than simply scaling up model size; see good examples in point cloud processing (Qi et al., 2017; Li et al., 2018; Fuchs et al., 2020; Chen et al., 2021), visual tasks (Cohen & Welling, 2016; Xu et al., 2023), graph neural networks (Veličković et al., 2017; Kanatsoulis & Ribeiro, 2024), physics and chemistry (Faber et al., 2016; Eismann et al., 2021; Kondor, 2025), reinforcement learning and decision making (Wang et al., 2022; Qin et al., 2022), amongst others. When learning these symmetries, classic methods must be shown numerous transformed (sometimes manually) data examples, such as rotated or translated images, and be enforced to make invariant/equivariant predictions, which is usually computationally expensive and requires vast amounts of data (Bronstein et al., 2021). Equivariant neural networks address these challenges through either new architectures or optimization methods to enforce symmetries (Cohen & Welling, 2016), in which a useful approach is to enforce each layer of the network to be equivariant.

Despite the impressive performance, it remains unclear whether enforcing equivariance comes at the cost of reduced expressive power. This paper investigates this trade-off by analyzing whether restricting the

*Both authors contributed equally.

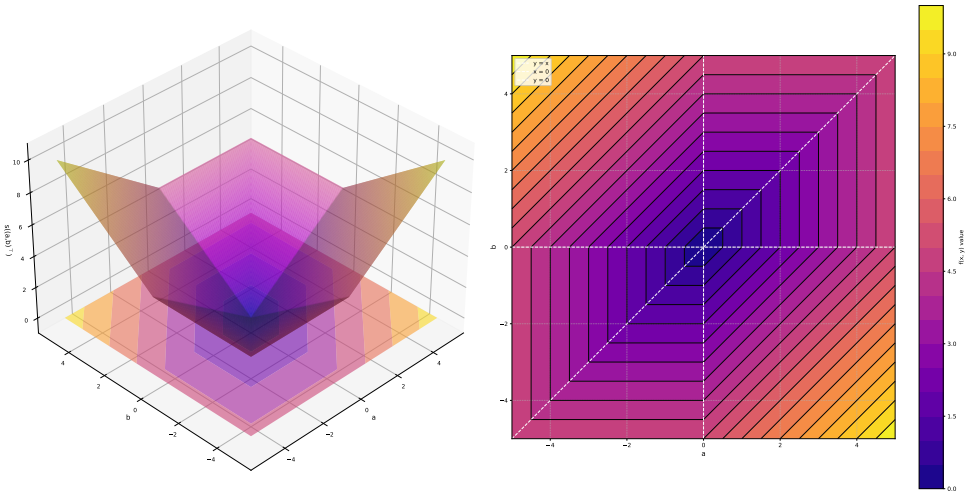

Figure 1: An equivariant function that satisfies $s((a,b)^\top) = s((b,a)^\top)$ and $s((a,b)^\top) = s((-a,-b)^\top)$. The left subfigure is a 3D plot, while the right subfigure is a 2D Contour map.

hypothesis space to equivariant models increases the expected risk when approximating an equivariant target function $s(x)$. Specifically, we examine whether the following inequalities hold:

$$\inf_{\theta\in\Theta} \mathsf{E}_{x\sim P}\left[\ \|F_\theta(x) - s(x)\|_2^2\right] < \inf_{\theta\in\Theta\cap\mathsf{GENs}} \mathsf{E}_{x\sim P}\left[\ \|F_\theta(x) - s(x)\|_2^2\right]$$
$$\text{and} \quad \inf_{\theta\in\Theta\cap\mathsf{GENs}} \mathsf{E}_{x\sim P}\left[\ \|F_\theta(x) - s(x)\|_2^2\right] < \inf_{\theta\in\Theta\cap\mathsf{LENs}} \mathsf{E}_{x\sim P}\left[\ \|F_\theta(x) - s(x)\|_2^2\right].$$

Here, Θ is the set of network parameters (usually L matrices) and F_θ is the neural network with parameter θ. We consider two types of equivariant networks: GENs (General Equivariant Networks), where the global mapping $x \to F_\theta(x)$ is required to be equivariant, and LENs (Layer-wise Equivariant Networks), a stricter subset where equivariance is enforced at every layer of the network. In practice, constructing LENs is a common approach to achieving GENs (He et al., 2021).

Boundary hyperplanes and channel vectors. We first investigate how these equivariance constraints (imposed at the level of input-output mapping) fundamentally restrict the architectural structure of neural networks. For two-layer ReLU networks, we focus on the geometric properties of the boundary hyperplanes and channel vectors of equivariant networks. We prove that GEN requires boundary hyperplanes to be symmetric. While for LENs, we demonstrate a constructive property: any LEN can be reformulated into another LEN that maintains the same model size but possesses a symmetric set of channel vectors. These geometric rigidities serve as the foundation for our subsequent analysis, enabling us to quantify how much model size is enough for approximating a target function.

Drawback of enforcing equivariance. Following the insight of the above analysis, we demonstrate a scenario where equivariance strictly hurts the expressive power. Specifically, we construct an invariant target function s on R^2 and establish two key findings: (1) for two-layer ReLU networks with a single neuron, general networks (GNs) achieve a strictly lower expected error compared to general equivariant networks (GENs), and (2) for two-layer ReLU networks that can arbitrarily approximate the target function, layer-wise equivariant networks (LENs) require a strictly larger model size than GENs.

Compensations via enlarging model size. Inversely, we prove that an increase in model size can fix the compromised expressive power. When the target function is invariant, and the data distribution is symmetric, both in terms of group G, enlarging the model size of any equivariant network by $|G|$ times can compensate for the expected loss over a 2-layer neural network. Moreover, a doubled model size allows LENs to represent all invariant functions of GENs. We further prove that the constructed LEN with $|G|$ times

enlarged model size surprisingly has a less complex hypothesis space, implying even improved generalizability. Together with the comparable expressive power, this result renders LENs a competitive solution.

To the best of our knowledge, this paper provides the first in-depth theoretical examination of the expressive power of layer-wise equivariant neural networks, demonstrating that enforcing equivariance has drawback in expressive power. We also show that an appropriately larger model size can compensate for the loss in expressive power and potentially enhance model generalisability, underlining the advantages of incorporating layer-wise equivariance into model architectures.

2 Related Works

Enforcing equivariance. Typical approaches for obtaining an equivariant neural network are (1) designing equivariant architectures (Cohen & Welling, 2017; Satorras et al., 2021; Trang et al., 2024) and (2) introducing equivariance constraints into the optimization (Winter et al., 2022; Tang, 2022; Pertigkiozoglou et al., 2024).

Equivariant neural architectures. Cohen & Welling (2016); Shin et al. (2016) design group convolutional neural networks in the group spaces as the equivariant variants of the convolutional neural networks, which is extended to the homogeneous spaces by Cohen et al. (2019). Similar approaches have also been employed in the design of other popular equivariant networks, such as graph neural networks (Klicpera et al., 2020; Aykent & Xia, 2025), transformers (Hutchinson et al., 2021; Islam et al., 2025), and diffusion models (Hoogeboom et al., 2022; Wan et al., 2025). Hao et al. (2025) leverage permutation equivariance in vision transformer for producing human-imperceptible, machine-recognizable images.

Symmetry-aware optimization. In this stream, equivariance constraints are introduced in optimization, as either regularisers or constraints, to ensure the equivariance, usually termed as steerable neural networks (Cohen & Welling, 2017). Other good examples are about enforcing the symmetry group S_N in steerable graph networks (Maron et al., 2018), the rotation and Euclidean groups $\{C_N; E_2\}$ in steerable convolutional neural networks (Weiler et al., 2018; Weiler & Cesa, 2019), and Lie symmetry in differential equations (Akhound-Sadegh et al., 2023; Jiang et al., 2025). A major obstacle is the high computational costs to compute the equivariant basis. Finzi et al. (2021) address this issues with an algorithm with polynomial computational complexity on the sum of the number of discrete generators and the dimension of the group, which divides the problem into some independent subproblems and adopts Krylov method to compute nullspaces.

Theoretical analysis. Theoretical studies have shown that equivariant neural networks have significant advantages, in terms of generalizability, convergence, and approximation. Sannai et al. (2021) prove improved generalization error bounds of equivariant models. Qin et al. (2022) leverage orbit averaging to construct a projection mapping from the original hypothesis space to the permutation-equivalent counterpart, underpinning reduced hypothesis complexity and better generalizability. Lawrence et al. (2022) prove that linear group convolutional neural networks trained by gradient descent for binary classification converge to solutions with low-rank Fourier matrix coefficients based on the results via implicit bias. Zaheer et al. (2017); Maron et al. (2019); Yarotsky (2021); Pacini et al. (2025) prove the universal approximation ability of equivariant models by various approximation techniques, while Ravanbakhsh (2020) utilize group averaging to obtain an equivariant approximator to prove the universal approximation ability. In addition, Elesedy & Zaidi (2021) prove that invariant/equivariant models have a smaller expected loss when the target function is invariant/equivariant. Azizian & marc lelarge (2021) study the expressive power of invariant/equivariant graph neural networks. On the other hand, Petrache & Trivedi (2023) focus on another approach that relaxes the strict constraints to partial and approximate equivariance, and provide bounds on the approximation error and generalization error. In contrast to the aforementioned studies, this paper focuses more on the geometric properties of ReLU networks and their expressive power. We prove that with the same model size, equivariance constraints impair expressive power. For the observed benefits of equivariance, we attribute them to the reduced hypothesis complexity, which holds even if the model size is larger.

3 Problem Settings

General networks (GNs). GNs refer to general neural networks without any design for realizing the equivariance. In this paper, we focus on two-layer networks, with a hidden layer of width m (also the number of neurons), representing a mapping from $\mathbb{R}^n \to \mathbb{R}^d$. The network is parameterized by θ, including the weight matrix $W^{(1)} \in \mathbb{R}^{m \times n}$ for the hidden layer, and the weight matrix $W^{(2)} \in \mathbb{R}^{d \times m}$ for an output layer. With the popular nonlinear activation ReLU $(\cdot)$, the GN maps any input $x \in \mathbb{R}^n$ to the output: $F(x) = W^{(2)}(W^{(1)}x)$. For the brevity, we denote by $w_i^\top$ the i-th column of $W^{(1)}$, by v_j the j-th row of $W^{(2)}$ and $W^{(1)}$, and collect these parameters (with m neurons) into a set θ_m. In this formulation, the output of a GN can be formulated as follows:

$$F(x) = W^{(2)}(W^{(1)}x) = \sum_{i=1}^{m} v_i (\langle w_i, x \rangle), \tag{1}$$

where $\langle w, x \rangle$ is the inner product $w^\top x$. We note that the total number of parameters in the GN is $nm + md$. As this count is proportional to the hidden layer size m, we use m to characterize the model size.

General equivariant networks (GENs). GENs represent a subset of GNs that constrain the output function F to be equivalent with respect to a given group representation. A group representation for a group G is formally de ned as a homomorphism $\rho : G \to GL(m)$, where GL(m) is the group of all invertible matrices on $\mathbb{R}^{m \times m}$. A function F is considered equivariant if, for given input and output representations ρ and ρ' respectively, it adheres to the condition $F(\rho_g x) = \rho'_g F(x)$ for all group elements $g \in G$. A GN whose function F satis es this property is categorised as a GEN. Speci cally, when $\rho' = \mathrm{id}$, the function F is called an invariant function and the networks are called invariant networks. In this paper, we assume that group G is nite and focus on invariant functions.

Layer-wise equivariant networks (LENs). LENs represent a subset of GNs that constrain each layer to be equivalent with respect to a given group representation. Speci cally, given the group representation ρ in the hidden layer, LENs require that for all $g \in G$,

$$\rho_g W^{(2)} = W^{(2)} \rho'_g; \quad \rho'_g = \rho_g; \quad \text{and} \quad \rho_g W^{(1)} = W^{(1)} \rho_g.$$

As each layer is constrained to be equivariant to maintain the symmetry in the data, the whole network is equivariant. Speci cally, for every group element $g \in G$, we have

$$\begin{aligned} F(\rho_g x) &= W^{(2)} W^{(1)} \rho_g x = W^{(2)} \rho'_g W^{(1)} x; \\ &= W^{(2)} \rho'_g W^{(1)} = \rho'_g W^{(2)} W^{(1)} = \rho'_g F(x). \end{aligned}$$

In practice, constructing LENs is a common approach to achieving GENs (He et al., 2021).

Expressive power. The expressive power, of GNs, GENs, and LENs is de ned as the minimum expected loss for approximating a target function s. Naturally, a smaller minimum expected loss means a better expressive power. Since GENs are in a subset of GNs, and LENs have intuitively more constraints, we have the following straightforward results: for any target function s, any input distribution $P \in \mathbb{R}^n$, and any number of neurons m,

$$\inf_{\theta \in \Theta_{\mathrm{GNs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2] \le \inf_{\theta \in \Theta_{\mathrm{GENs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2] \le \inf_{\theta \in \Theta_{\mathrm{LENs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2];$$

where θ represents the parameter of a two-layer ReLU network ($\theta = (W^{(2)}, W^{(1)})$), and F_θ is the forward function given parameter θ, and Θ_m is the parameter set of m-hidden-layer networks. We reuse the notations GNs (also for GENs and LENs) to denote the respective parameter sets that makes F_θ a GN (also for GEN and LEN). This paper investigates whether equivariance constrains the expressive power of two-layer networks. Speci cally, we study whether there exists some equivariant function s and m such that

$$\begin{aligned} &\inf_{\theta \in \Theta_{\mathrm{GNs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2] < \inf_{\theta \in \Theta_{\mathrm{GENs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2]; \text{ and} \\ &\inf_{\theta \in \Theta_{\mathrm{GENs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2] < \inf_{\theta \in \Theta_{\mathrm{LENs}}} \mathbb{E}_{x \sim P}[\|F_\theta(x) - s(x)\|_2^2]. \end{aligned}$$