Symmetry Breaking and Equivariant Neural Networks

Sékou-Oumar Kaba   KABASEKO@MILA.QUEBEC
Mila - Quebec AI Institute
School of Computer Science, McGill University

Siamak Ravanbakhsh   SIAMAK@CS.MCGILL.CA
Mila - Quebec AI Institute
School of Computer Science, McGill University

Editors: Sophia Sanborn, Christian Shewmake, Simone Azeglio, Nina Miolane

Abstract

Using symmetry as an inductive bias in deep learning has been proven to be a principled approach for sample-efficient model design. However, the relationship between symmetry and the imperative for equivariance in neural networks is not always obvious. Here, we analyze a key limitation that arises in equivariant functions: their incapacity to break symmetry at the level of individual data samples. In response, we introduce a novel notion of 'relaxed equivariance' that circumvents this limitation. We further demonstrate how to incorporate this relaxation into equivariant multilayer perceptrons (E-MLPs), offering an alternative to the noise-injection method. The relevance of symmetry breaking is then discussed in various application domains: physics, graph representation learning, combinatorial optimization and equivariant decoding.

Keywords: deep learning, invariance, equivariance, symmetry breaking, graph representation learning, physics

1. Introduction

The notion of symmetry is of fundamental importance across the sciences, mathematics, and more recently in machine learning. It captures the idea that an object is essentially the same after some transformation is applied to it (Weyl, 1952). Using symmetry as an inductive bias in machine learning has emerged as a powerful idea, with important conceptual and practical breakthroughs (Bronstein et al., 2021).

The common intuition is that symmetry in the *data distribution* should naturally lead to equivariance constraints on learned functions. However, even in symmetric domains, it appears that equivariant functions have an important limitation: the inability to break symmetry at the level of *data samples*. The classical example of symmetry breaking appears in physical phase transitions. From an initially symmetric state, an asymmetric state is observed (see Section 1). As we will see and as discussed by Smidt et al. (2021), equivariant neural networks are unable to model these phenomena. Getting rid of equivariance altogether would be an unsatisfactory solution, as it is still necessary to account for the symmetry of physical laws.

In this theory-oriented extended abstract, we give a precise characterization of this problem and argue that it is not limited to applications in physics. We show that a wide range of learning tasks require symmetry breaking and that equivariance is therefore fundamentally too constraining. We introduce a relaxation of equivariance that allows to deal with this

issue. We then show how to build equivariant multilayer perceptrons (Shawe-Taylor, 1989; Ravanbakhsh et al., 2017; Finzi et al., 2021) that can break symmetry. Finally, we propose avenues for future works and practical applications of our framework.

We introduce some mathematical background and notations used in the rest of the paper in Appendix A.

2. Equivariance Preserves Symmetry

It is known that equivariant functions preserve the symmetry of their input. One of the earliest versions of this statement is due to Curie (1894): "*the symmetries of the causes are to be found in the effects*". Chalmers (1970) provided a more mathematical version of this statement, with effects (observed state) being the result of equivariant physical laws acting on causes (initial state). The general idea is captured by the following proposition.

Proposition 1 (Curie's Principle) *Let ϕ be an equivariant function and $G_{\mathbf{x}}$ denote the stabilizer subgroup of $\mathbf{x}$. Then,*

$$G_{\phi(\mathbf{x})} \geq G_{\mathbf{x}}, \forall \mathbf{x} \in \mathcal{X}.$$

The proof follows in Appendix C.1. This can also be said differently in terms of orbit types (see definition in Appendix A). When the equivariant function is seen as acting on orbits, we must have $\phi(G \cdot \mathbf{x}) \lesssim G \cdot \mathbf{x}$. An equivariant function therefore cannot map an orbit of type $[G_{\mathbf{x}}]$ to an orbit of type that is not coarser than $[G_{\mathbf{x}}]$.

For continuous functions, a version of this result holds when inputs are approximately symmetric. In this case, the inability to break the symmetry for symmetric inputs translates to more difficulty in breaking it for approximately symmetric inputs.

Proposition 2 *Let ϕ be equivariant and Lipschitz, with constant k and $\|\cdot\|$ denote the induced norm. Then,*

$$\|g \cdot \phi(\mathbf{x}) - \phi(\mathbf{x})\| \leq k\|g \cdot \mathbf{x} - \mathbf{x}\|, \forall g, \mathbf{x} \in G \times \mathcal{X}.$$

The proof follows in Appendix C.2. If an input is close to its transformed version, the images under a continuous equivariant function also have to be close.

Finally, we highlight an important fact regarding symmetric inputs of finite groups.

Proposition 3 *Let $\mathcal{X} = \mathbb{R}^n$ and $\rho : G \to GL(\mathcal{X})$ be any non-trivial linear group action of a finite group with faithful representation. Then, the set of symmetric inputs $S = \{\mathbf{x} \in \mathcal{X} \mid G_{\mathbf{x}} \neq \{e\}\}$ is of measure zero with respect to the Lebesgue measure.*

The proof follows in Appendix C.3. This captures many groups of interest in machine learning. Symmetric inputs are therefore in some sense *rare*. At first glance, this could suggest that the Curie principle (Proposition 1) is hardly relevant since the cases in which it would apply are improbable. Things are however not so simple. First, in many domains, such as graphs, the set of actual inputs is discrete, e.g. $\mathcal{X} = V^n$, where V is a finite set. In this case, Proposition 3 does not apply. Second, there could be a significant bias towards

symmetric inputs in the data, as these data points often have special properties that make them more common. This is for example often the case in physics. Third, non-injective activation functions, like ReLUs (Nair and Hinton, 2010), can make symmetric activations much more likely in the intermediary layers of a neural network by zeroing out entries. It is therefore important to handle symmetric inputs beyond the constraints imposed by equivariance, as we explain in the next section.

3. Relaxed Equivariance

A version of equivariance that allows breaking the symmetry of inputs and mapping to arbitrary orbit types is necessary. Some applications are detailed in Section 5. We note that the appropriate notion was introduced by (Kaba et al., 2023) for canonicalization, a problem requiring symmetry breaking. However, their definition applies more generally.

Definition 4 (Relaxed equivariance) *Given group actions on $\mathcal{X}$ and $\mathcal{Y}$, $\phi : \mathcal{X} \to \mathcal{Y}$ satisfies relaxed equivariance if $\forall g_1, \mathbf{x} \in G \times X$, there exists $g_2 \in g_1 G_{\mathbf{x}}$ such that*

$$\phi\left(g_1 \cdot \mathbf{x}\right) = g_2 \cdot \phi\left(\mathbf{x}\right). \tag{1}$$

The motivation for relaxed equivariance being the correct way to account for symmetry breaking is as follows. First, it captures the idea of symmetry in the task, meaning that the output of the function is predictable under transformation of the input, *up to* meaningless stabilizing transformations since $\phi\left(g_1 \cdot \mathbf{x}\right) = g_1 \cdot g_{\mathbf{x}} \cdot \phi\left(\mathbf{x}\right)$, with $g_{\mathbf{x}} \in G_{\mathbf{x}}$. Second, the output does not need to maintain all the symmetries of the input. To see this, notice that for $g_1 \in G_{\mathbf{x}}$, one possibility allowed by relaxed equivariance is $g_{\mathbf{x}} = g_1^{-1}$. In this case, we obtain $\phi\left(g_1 \cdot \mathbf{x}\right) = \phi\left(\mathbf{x}\right)$, which by contrast to what we have with equivariance (see Appendix C.1), does not impose any constraints on the stabilizer of the output.

In Appendix B, we further justify how relaxed equivariance naturally appears in machine learning from first principles.

4. Breaking Symmetry in Equivariant Multilayer Perceptrons

We now investigate how to build relaxed equivariance into neural networks instead of equivariance. One seemingly ad-hoc solution is sometimes adopted to deal with symmetry breaking, for example by Liu et al. (2019) and Locatello et al. (2020) for graph and set generation. It simply consists of adding noise to the input to break the symmetry and then using an equivariant neural network. Proposition 3 confirms that this procedure has some justification. The input is almost surely mapped to a regular orbit. Then, the equivariant neural network can map the noisy input to an orbit of arbitrary type. However, there are at least two downsides to this approach. First, relaxed equivariance is only respected in expectation, similarly to equivariance when adding noise to data. Second, if the subsequent equivariant neural network is continuous, Proposition 2 indicates that a significant amount of noise will be required to properly break the symmetry, which might hurt generalization.

To circumvent these issues, we provide an adaptation of equivariant multilayer perceptrons (E-MLPs) that can handle symmetry breaking (Shawe-Taylor, 1989; Ravanbakhsh et al., 2017; Finzi et al., 2021). E-MLPs provide a standard method to build equivariant

neural networks (Bronstein et al., 2021) and consist of stacking linear equivariant layers with point-wise non-linear functions.

Linear layers with relaxed equivariance can be constructed using the following result:

Theorem 5 *Let* $\mathbf{W} \in \mathbb{R}^{m \times n}$ *be a weight matrix,* $\mathbf{x} \in \mathbb{R}^n$ *be an input with stabilizer* $G_\mathbf{x}$, $\mathcal{X} = G \cdot \mathbf{x}$ *be the orbit of* $\mathbf{x}$ *and* $\rho : G \to GL(\mathbb{R}^n)$, $\rho' : G \to GL(\mathbb{R}^m)$ *be group representations of* G *on the input and output respectively. If, for all left cosets* $C \in G/G_\mathbf{x}$

$$\rho'(g)\,\mathbf{W}\mathbf{x} = \mathbf{W}\rho(g)\,\mathbf{x}, \tag{2}$$

where $g \in C$ *is an arbitrary coset representative, then the map* $\boldsymbol{\phi} : \mathcal{X} \to \mathbb{R}^m; \mathbf{x} \mapsto \mathbf{W}\mathbf{x}$ *satisfies relaxed equivariance.*

The proof and some discussion follow in Appendix C.5. Additionally, for permutation groups standard point-wise activation functions can be used, thanks to the fact that they satisfy relaxed equivariance (Appendix D.1), and that relaxed equivariance is compatible with composition (Appendix D.2).

5. Applications

Our analysis provides a general framework for symmetry breaking in deep learning and applies to multiple domains. We give a few examples thereafter of domains in which we think symmetry breaking analysis could be an exciting future direction.

Physics modelling Symmetry breaking was first described in physics. Being able to break symmetry is important to describe phase transitions, notably in condensed matter systems (Kaba and Ravanbakhsh, 2022) and bifurcations in dynamical systems (Golubitsky and Stewart, 2002).

Graph representation learning Learned node representations in a graph will carry the same symmetry as the graph itself. This is often not a requirement and can be detrimental. The simplest example is the task of predicting edges using node representations: nodes within the same orbit of the automorphism group must be assigned the same neighbourhoods (Satorras et al., 2021; Lim et al., 2023).

Combinatorial optimization Machine learning can be used for combinatorial optimization problems (Bengio et al., 2021). With discrete structures, if we want to predict single solutions, it is necessary to handle degeneracies caused by symmetry – that is we need to break the symmetry to identify a single solution.

Equivariant decoding Designing a decoder from an invariant latent space to a space on which a group acts non-trivially is not a well-defined problem (Severo et al., 2021; Vignac and Frossard, 2022; Zhang et al., 2022). This can, however, be seen as an instance of symmetry breaking. This is related to the discussion of Appendix B.

6. Conclusion

In this paper, we have analyzed a fundamental limitation of equivariant functions in handling symmetry breaking. We have shown that it is important to account for it in multiple

applications in machine learning by relaxing the equivariance constraint. We have finally provided a way to adapt E-MLPs to satisfy the relaxed version equivariance instead of the standard one. We hope this constitutes a first effort to better understand symmetry breaking in machine learning. Many avenues are still left to explore for the extension of this work. First, experimental testing of our claims in different domains is necessary. Second, the constraint stated in Theorem 5 could be costly to solve for large groups; making it scale sublinearly in group size would be desirable. Finally, alternative ways to achieve relaxed equivariance could be explored, notably a probabilistic approach where the symmetry equivalent images are sampled instead of being deterministically computed by the network.

Acknowledgments

References

Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: a methodological tour d'horizon. *European Journal of Operational Research*, 290(2):405–421, 2021.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Velicković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Alan F Chalmers. Curie's principle. *The British Journal for the Philosophy of Science*, 21 (2):133–148, 1970.

Pierre Curie. Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique. *Journal de physique théorique et appliquée*, 3 (1):393–415, 1894.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International conference on machine learning*, pages 3318–3328. PMLR, 2021.

Martin Golubitsky and Ian Stewart. *The symmetry perspective: from equilibrium to chaos in phase space and physical space*, volume 200. Springer Science & Business Media, 2002.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

Sékou-Oumar Kaba and Siamak Ravanbakhsh. Equivariant networks for crystal structures. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=0Dh8dz4snu.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pages 15546–15566. PMLR, 2023.

Derek Lim, Joshua Robinson, Stefanie Jegelka, Yaron Lipman, and Haggai Maron. Expressive sign equivariant networks for spectral geometric learning. In *ICLR 2023 Workshop on Physics for Machine Learning*, 2023.

Jenny Liu, Aviral Kumar, Jimmy Ba, Jamie Kiros, and Kevin Swersky. Graph normalizing flows. *Advances in Neural Information Processing Systems*, 32, 2019.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in Neural Information Processing Systems*, 33:11525–11538, 2020.

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

Charles C Pinter. *A book of abstract algebra*. Courier Corporation, 2010.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. In *International Conference on Machine Learning*, pages 2892–2901. PMLR, 2017.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

Daniel Severo, James Townsend, Ashish J Khisti, Alireza Makhzani, and Karen Ullrich. Your dataset is a multiset and you should compress it like one. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

J. Shawe-Taylor. Building symmetries into feedforward networks. In *1989 First IEE International Conference on Artificial Neural Networks, (Conf. Publ. No. 313)*, pages 158–162, 1989.

Tess E. Smidt, Mario Geiger, and Benjamin Kurt Miller. Finding symmetry breaking order parameters with euclidean neural networks. *Phys. Rev. Research*, 3:L012002, Jan 2021. doi: 10.1103/PhysRevResearch.3.L012002. URL https://link.aps.org/doi/10.1103/PhysRevResearch.3.L012002.

Clement Vignac and Pascal Frossard. Top-n: Equivariant set and graph generation without exchangeability. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=-Gk_IPJWvk.

Hermann Weyl. Symmetry. In *Symmetry*. Princeton University Press, 1952.

Yan Zhang, David W Zhang, Simon Lacoste-Julien, Gertjan J. Burghouts, and Cees G. M. Snoek. Multiset-equivariant set prediction with approximate implicit differentiation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=5K7RRqZEjoS.

Appendix A. Background

In the following section, we introduce some useful notions on group actions and equivariant functions. The results we refer to can be found in elementary textbook on group theory, for example Pinter (2010).

Groups actions Given a group G and a set $\mathcal{X}$, a (left) group action is a function $a : G \times \mathcal{X} \to \mathcal{X}$, such that

$$a\left(e, \mathbf{x}\right) = \mathbf{x} \qquad\qquad a\left(g, \left(a\left(h, \mathbf{x}\right)\right)\right) = a\left(gh, \mathbf{x}\right).$$

We will use the shorthand notation $a\left(g, \mathbf{x}\right) = g \cdot \mathbf{x}$ The nature of the action $\cdot$ will be clear from context.

A group action is *transitive* if for any $\mathbf{x}, \mathbf{x}'$, there exists a g such that $g \cdot \mathbf{x} = \mathbf{x}'$. This means that any elements can be mapped into another by the group action.

We will be mostly interested in linear group actions, for which $g \cdot \mathbf{x} = \rho\left(g\right)\mathbf{x}$ and $\rho : G \to GL\left(\mathcal{X}\right)$ is a group homomorphism called a *representation* of the group. The representation is *faithful* if ρ is injective.

Orbit types The *orbit* of an element $\mathbf{x}$, is defined as $G \cdot \mathbf{x} \equiv \{g \cdot \mathbf{x} \mid g \in G\}$. It is the set of elements to which $\mathbf{x}$ can be mapped to by the group action. The set of orbits under the group action, denoted $\mathcal{X}/G$ forms a partition of $\mathcal{X}$. The group action is transitive if and only if has only one orbit.

The stabilizer of an element $\mathbf{x}$, is defined as $G_{\mathbf{x}} \equiv \{g \in G \mid g \cdot \mathbf{x} = \mathbf{x}\}$. $\mathbf{x}$ is called a *fixed point* or *invariant* if $G_{\mathbf{x}} = G$. If $G_{\mathbf{x}} = \{e\}$, the action is said to be *regular* on $G \cdot \mathbf{x}$.

It can be shown that $G_{g \cdot \mathbf{x}} = g G_{\mathbf{x}} g^{-1}$, e.g. the stabilizers of elements in the same orbit are conjugate. We can therefore associate to each orbit a conjugacy class of a subgroup of G, which are the stabilizers on the orbit. Let H be the stabilizer of some element of $G \cdot \mathbf{x}$ and $[H] \equiv \{gHg^{-1} \mid g \in G\}$ be the conjugacy class of H. Then, the orbit $G \cdot \mathbf{x}$ is said to be of type $[H]$.

It can additionally be shown that the group action on an orbit of type $[H]$ is isomorphic to the group action on the cosets G/H defined by

$$b : G \times G/H \to G/H; \qquad\qquad g_1, g_2 H \mapsto (g_1 g_2) H \qquad\qquad (3)$$

Group actions on orbits of the same type are therefore isomorphic. We will use the symbol $\simeq$ to denote equivalence. We also introduce an order relation between orbits based on that equivalence. If $G \cdot \mathbf{x} \simeq [H_1]$ and $H_1 \geq H_2$, we will say that $G \cdot \mathbf{x} \lesssim [H_2]$. This is motivated by the orbit-stabilizer theorem: if an orbit has a bigger stabilizer, then it must be smaller in size.

Orbit types allow to classify group actions. Since the orbits induce a partition $\mathcal{X}$, it is natural to decompose $\mathcal{X}$ into orbits, with

$$\mathcal{X} = \bigcup_{\mathbf{x} \in c(\mathcal{X}/G)} G \cdot \mathbf{x} \cong \bigcup_{\mathbf{x} \in c(\mathcal{X}/G)} [G_{\mathbf{x}}] \tag{4}$$

where c is a choice function over the partitions.

The kernel of a group action is defined as $\text{Ker}\,(a) \equiv \{g \in G \mid g \cdot \mathbf{x} = \mathbf{x}, \ \forall \ \mathbf{x} \in \mathcal{X}\}$. It can be shown that $\text{Ker}\,(a) = \bigcap_{\mathbf{x} \in \mathcal{X}} G_{\mathbf{x}}$. If an orbit is of type $[H]$, where H is a normal subgroup, then $\text{Ker}\,(a) = H$.

Equivariant functions   Given (possibly differentt) group actions on $\mathcal{X}$ and $\mathcal{Y}$, an equivariant function is a function $\phi : \mathcal{X} \to \mathcal{Y}$ such that

$$\phi\,(g \cdot \mathbf{x}) = g \cdot \phi\,(\mathbf{x}) \tag{5}$$

An equivariant function can therefore be seen as a homomorphism between group actions.

It follows immediately that equivariance preserves orbits

$$\phi\,(G \cdot \mathbf{x}) = G \cdot \phi\,(\mathbf{x}) \tag{6}$$

We therefore naturally obtain that ϕ induces a mapping between orbits, $\phi : \mathcal{X}/G \to \mathcal{Y}/G$.

If K is the kernel of the group action on $\mathcal{Y}$, the function is considered K invariant and G/K equivariant. In particular, when $K = G$, the function is simply called invariant.

Appendix B. First-principle Derivation of Relaxed Equivariance

We are in general interested in learning tasks for which the underlying distribution possesses some symmetry. For predictive modelling, given some group actions on $\mathcal{X}$ and $\mathcal{Y}$, that means that underlying conditional distribution satisfies $p\,(g \cdot \mathbf{y} | g \cdot \mathbf{x}) = p\,(\mathbf{y}|\mathbf{x}) \ \forall g \in G$. This is similar when modelling data conditioned on a latent variable with $p\,(g \cdot \mathbf{x} | g \cdot \mathbf{z}) = p\,(\mathbf{x}|\mathbf{z}) \ \forall g \in G$. When we wish for the model to approximate the full distribution on $\mathcal{Y}$ (typically when $|\mathcal{Y}|$ is finite and small), equivariance with the action defined on functions follows straightforwardly. In that case, we assume $\phi : \mathcal{X} \to \mathbb{P}\,(\mathcal{Y}), \ \mathbf{x} \mapsto \phi\,[\mathbf{y}]\,(\mathbf{x})$, where $\mathbb{P}\,(\mathcal{Y})$ is the set of probability distributions on $\mathcal{Y}$ and obtain

$$\phi\,[\mathbf{y}]\,(g \cdot \mathbf{x}) = \phi\,\left[g^{-1} \cdot \mathbf{y}\right]\,(\mathbf{x}) = (g \cdot \phi)\,[\mathbf{y}]\,(\mathbf{x}) \tag{7}$$

However, in many situations, we wish to obtain a *deterministic model* giving an output that maximizes the probability, rather than modelling the full distribution; e.g., Maximum a Posteriori instead of the full posterior Hastie et al. (2009) (note that a similar argument applies when trying to approximate the distribution by a simpler one). In this case, we define $\phi : \mathcal{X} \to \mathcal{Y}$ as

$$\phi\,(\mathbf{x}) = c \left(\underset{\mathbf{y} \in \mathcal{Y}}{\arg\max}\, p\,(\mathbf{y}|\mathbf{x}) \right) \tag{8}$$

where the $\arg\max$ is a set since the maximum may not be unique and $c : 2^{\mathcal{Y}} \to \mathcal{Y}$ is a choice function that selects a unique element.

We show in Appendix C.4, that if the distribution is symmetric under some group action, then $\arg\max_{\mathbf{y}\in\mathcal{Y}} p\left(\mathbf{y}|\mathbf{x}\right)$ must be a union of orbits of the stabilizer of $\mathbf{x}$ when acting on $\mathcal{Y}$. This is simply because, then, some probabilities are the same by symmetry.

We now assume that $\arg\max_{\mathbf{y}\in\mathcal{Y}} p\left(\mathbf{y}|\mathbf{x}\right)$ is a unique orbit. In a sense, this amounts to the idea that all the symmetry of the model is completely captured by the transformation group G. We can then prove the following theorem:

Theorem 6 *Let ϕ be defined by Equation (8). If p is symmetric under some action of G and the set $\arg\max_{\mathbf{y}\in\mathcal{Y}} p\left(\mathbf{y}|\mathbf{x}\right)$ is a unique orbit, then ϕ satisfies the relaxed equivariance condition.*

The proof is given in Appendix C.4. Relaxed equivariance therefore naturally arises as a requirement for deterministic models under symmetric distributions. The same applies when ϕ is a function that generates samples from a latent variable when the underlying conditional distribution $p\left(\mathbf{x}|\mathbf{z}\right)$ is symmetric.

Appendix C. Proofs

C.1. Proposition 1

Proof For any $\mathbf{x}\in\mathcal{X}$ and $g\in G_{\mathbf{x}}$, we have

$$\phi\left(g\cdot\mathbf{x}\right)=\phi\left(\mathbf{x}\right). \tag{9}$$

From equivariance of ϕ, we also have

$$\phi\left(g\cdot\mathbf{x}\right)=g\cdot\phi\left(\mathbf{x}\right). \tag{10}$$

Thus,

$$g\cdot\phi\left(\mathbf{x}\right)=\phi\left(\mathbf{x}\right) \tag{11}$$

The stabilizer of $\phi\left(\mathbf{x}\right)$ is therefore at least $G_{\mathbf{x}}$, which completes the proof. ∎

C.2. Proposition 2

Proof If ϕ is Lipschitz with constant k, we have

$$\|\phi\left(g\cdot\mathbf{x}\right)-\phi\left(\mathbf{x}\right)\|\leq k\|g\cdot\mathbf{x}-\mathbf{x}\|,\forall g,\mathbf{x}\in G\times\mathcal{X}. \tag{12}$$

From equivariance of ϕ, we find

$$\|g\cdot\phi\left(\mathbf{x}\right)-\phi\left(\mathbf{x}\right)\|\leq k\|g\cdot\mathbf{x}-\mathbf{x}\|,\forall g,\mathbf{x}\in G\times\mathcal{X}. \tag{13}$$

which completes the proof. ∎

C.3. Proposition 3

Proof

The set S is equal to $\bigcup_{g \in G/\{e\}} \{x \in X \mid g \cdot x = x\}$. We will show that for each $g \in G/\{e\}$, the set of elements of $\mathcal{X}$ stabilized by g is of measure zero. Since the union is over a finite set, S will therefore also be of measure zero.

The set of elements stabilized by g is given by the solutions of the equation $\rho(g)\mathbf{x} = \mathbf{x}$. The stabilizer is therefore the eigenspace of $\rho(g)$ with eigenvalues 1. If ρ is a faithful representation, then for any $g \neq e$, $\rho(g) \neq I$. However, for any linear operator other than I, the dimension of eigenspaces with eigenvalue 1, if they exist, must be $d < n$. But, any subspace of $\mathbb{R}^n$ of dimension $d < n$ has measure zero with respect to the Lebesgue measure. Therefore, the set of elements stabilized by any $g \neq e$ is of measure zero.

This completes the proof. ■

C.4. Theorem 6

We introduce the following lemmas

Lemma 7 *Let $p(\mathbf{y} \mid \mathbf{x}) = p(g \cdot \mathbf{y} \mid g \cdot \mathbf{x})$ for all $g \in G$. Then,*

$$G_{\mathbf{x}} \cdot \left(\arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \right) = \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \tag{14}$$

Proof From the symmetry of p, and based on the definition of the stabilizer, we have for all $g_{\mathbf{x}} \in G_{\mathbf{x}}$

$$p(\mathbf{y} \mid \mathbf{x}) = p(g_{\mathbf{x}} \cdot \mathbf{y} \mid g_{\mathbf{x}} \cdot \mathbf{x}) \tag{15}$$

$$p(\mathbf{y} \mid \mathbf{x}) = p(g_{\mathbf{x}} \cdot \mathbf{y} \mid \mathbf{x}) \tag{16}$$

Therefore,

$$\mathbf{y}^* \in \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \implies g_{\mathbf{x}}^{-1} \cdot \mathbf{y}^* \in \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \tag{17}$$

which concludes the proof. ■

Lemma 8 *Let $p(\mathbf{y} \mid \mathbf{x}) = p(g \cdot \mathbf{y} \mid g \cdot \mathbf{x})$ for all $g \in G$. Then,*

$$\left(\arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|g \cdot \mathbf{x}) \right) = g \cdot \left(\arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \right) \tag{18}$$

Proof From the symmetry p, we have for all $g \in G$

$$p(\mathbf{y} \mid g \cdot \mathbf{x}) = p(g^{-1} \cdot \mathbf{y} \mid \mathbf{x}) \tag{19}$$

Therefore,

$$\mathbf{y}^* \in \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|g \cdot \mathbf{x}) \implies g \cdot \mathbf{y}^* \in \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \tag{20}$$

which concludes the proof. ∎

We now provide the proof of Theorem 6.

Proof We have

$$\phi(\mathbf{x}) = c\left(\arg\max_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|\mathbf{x})\right) \tag{21}$$

Using Lemma 7, we therefore have

$$\phi(\mathbf{x}) \in G_\mathbf{x} \cdot \left(\arg\max_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|\mathbf{x})\right) \tag{22}$$

$$\phi(g\cdot\mathbf{x}) \in G_\mathbf{x} \cdot \left(\arg\max_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|g\cdot\mathbf{x})\right) \tag{23}$$

Using Lemma 8, we obtain

$$\phi(g\cdot\mathbf{x}) \in g\cdot G_\mathbf{x} \cdot \left(\arg\max_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|\mathbf{x})\right) \tag{24}$$

Using the assumption that the arg max is only one orbit, we have

$$\phi(g\cdot\mathbf{x}) \in g\cdot G_\mathbf{x} \cdot c\left(\arg\max_{\mathbf{y}\in\mathcal{Y}} p(\mathbf{y}|\mathbf{x})\right) \tag{25}$$

$$\phi(g\cdot\mathbf{x}) \in g\cdot G_\mathbf{x} \cdot \phi(\mathbf{x}) \tag{26}$$

This is equivalent to saying that there exists a $g_2 \in g\cdot G_\mathbf{x}$ such that

$$\phi(g\cdot\mathbf{x}) = g_2 \cdot \phi(\mathbf{x}) \tag{27}$$

which is the relaxed equivariance condition. ∎

C.5. Theorem 5

We first introduce the following lemma.

Lemma 9 *Consider group actions by G on $\mathcal{X}$ and $\mathcal{Y}$, $\mathbf{x} \in \mathcal{X}$ and $\phi : \mathcal{X} \to \mathcal{Y}$. If $\forall g_1 \in G$ there exists a $g_2 \in g_1 G_\mathbf{x}$ such that*

$$\phi(g_1 \cdot \mathbf{x}) = g_2 \cdot \phi(\mathbf{x}), \tag{28}$$

then $\forall g_1, \mathbf{y} \in G \times G \cdot \mathbf{x}$, there exists a $g_2 \in g_1 G_\mathbf{y}$ such that

$$\phi(g_1 \cdot \mathbf{y}) = g_2 \cdot \phi(\mathbf{y}). \tag{29}$$

Proof Any $\mathbf{y} \in G \cdot \mathbf{x}$ can be written as $\mathbf{y} = g \cdot \mathbf{x}$ for some $g \in G$. We have

$$\phi\left(g_1 \cdot \mathbf{y}\right) = \phi\left(g_1 \cdot g \cdot \mathbf{x}\right). \tag{30}$$

From Equation (28), we have

$$\phi\left(g_1 \cdot \mathbf{y}\right) = g_1 \cdot g \cdot g_\mathbf{x} \cdot \phi\left(\mathbf{x}\right), \tag{31}$$

for some $g_\mathbf{x} \in G_\mathbf{x}$. From Equation (28), we also find

$$\phi\left(g \cdot \mathbf{x}\right) = g \cdot g'_\mathbf{x} \cdot \phi\left(\mathbf{x}\right), \tag{32}$$

for some $g'_\mathbf{x} \in G_\mathbf{x}$. Therefore,

$$g'^{-1}_\mathbf{x} \cdot g^{-1} \cdot \phi\left(g \cdot \mathbf{x}\right) = \phi\left(\mathbf{x}\right). \tag{33}$$

Replacing in 31, we obtain

$$\phi\left(g_1 \cdot \mathbf{y}\right) = g_1 \cdot g \cdot g_\mathbf{x} \cdot g'^{-1}_\mathbf{x} \cdot g^{-1} \cdot \phi\left(\mathbf{y}\right). \tag{34}$$

Since we have $g_\mathbf{x} \cdot g'^{-1}_\mathbf{x} \in G_\mathbf{x}$ and $g \cdot g_\mathbf{x} \cdot g^{-1} \in G_\mathbf{y} \forall g_\mathbf{x} \in G_\mathbf{x}$, we have

$$\phi\left(g_1 \cdot \mathbf{y}\right) = g_1 \cdot g_\mathbf{y} \cdot \phi\left(\mathbf{y}\right), \tag{35}$$

where $g_\mathbf{y} \in G_y$. This completes the proof. ■

We now prove Theorem 5.

Proof Consider the map $\phi\left(\mathbf{x}\right) = \mathbf{W}\mathbf{x}$. Relaxed equivariance for ϕ implies that for all $g_1 \in G$ and for all $x \in \mathbb{R}^n$, there exists a $g_2 \in g_1 G_\mathbf{x}$ such that the constraint

$$\rho'\left(g_2\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_1\right)\mathbf{x} \tag{36}$$

is satisfied. Consider all g_1 that are elements of the same coset of the stabilizer, e.g. the set $g_1 G_\mathbf{x}$. For all these elements, the constraint Equation (36) can be satisfied with the same element g_2. We can therefore have for all these elements,

$$\rho'\left(g_2\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_2 \cdot g_\mathbf{x}\right)\mathbf{x}, \tag{37}$$

where $g_\mathbf{x} \in G_\mathbf{x}$ and a unique g_2 chosen arbitrarily in $g_1 G_\mathbf{x}$. By definition of the stabilizer, we have

$$\rho'\left(g_2\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_2\right)\rho\left(g_\mathbf{x}\right)\mathbf{x} \tag{38}$$

$$\rho'\left(g_2\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_2\right)\mathbf{x} \tag{39}$$

Therefore, if for all cosets in $G/G_\mathbf{x}$, Equation (39) is satisfied with an arbitrary representative, Equation (36) is satisfied for all $g_1 \in G$. Using Lemma 9, this implies that it is also satisfied for any $\mathbf{y} \in G \cdot \mathbf{x}$, which completes the proof.

■

For the coset containing the identity element, the representative can selected as the identity itself, such that there is no constraint. This therefore results in $|G|/|G_{\mathbf{x}}| - 1$ constraints.

Note that contrarily to standard equivariance constraints like in (Finzi et al., 2021), it does not follow from these constraints that if

$$\rho'\left(g_2\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_2\right)\mathbf{x} \tag{40}$$

$$\rho'\left(g_3\right)\mathbf{W}\mathbf{x} = \mathbf{W}\rho\left(g_3\right)\mathbf{x} \tag{41}$$

a similar constraint is also satisfied for $g_2 \cdot g_3$. This is because of the explicit multiplication by $\mathbf{x}$ that cannot be eliminated. It is therefore not possible to straightforwardly reduce the constraints to a set of generators.

Appendix D. Properties of Relaxed Equivariance

D.1. Equivariant functions

This property is trivially satisified, but it is still useful to formulate it explicitly.

Proposition 10 *Let ϕ be equivariant. Then, ϕ satisfies relaxed equivariance.*

Proof If ϕ is equivariant:

$$\phi\left(g \cdot \mathbf{x}\right) = g \cdot \phi\left(\mathbf{x}\right) \tag{42}$$

Since $g \in gG_{\mathbf{x}}$, ϕ satisfies the relaxed equivariance condition. ∎

D.2. Composition

Proposition 11 *Let $\phi_1 : \mathcal{X} \to \mathcal{Y}$ and $\phi_2 : \mathcal{Y} \to \mathcal{Z}$ satisfy relaxed equivariance. Then $\phi_2 \circ \phi_1$ satisfies relaxed equivariance.*

Proof We have

$$\phi_2\left(\phi_1\left(g_1 \cdot \mathbf{x}\right)\right) = \phi_2\left(g_2 \cdot \phi_1\left(\mathbf{x}\right)\right) \tag{43}$$

where $g_2 \in g_1 G_{\mathbf{x}}$. Then,

$$\phi_2\left(g_2 \cdot \phi_1\left(\mathbf{x}\right)\right) = g_3 \cdot \phi_2\left(\phi_1\left(\mathbf{x}\right)\right) \tag{44}$$

where $g_3 \in g_2 G_{\mathbf{x}}$. Since $g_2 G_{\mathbf{x}} = g_1 G_{\mathbf{x}}$, we have $g_3 \in g_1 G_{\mathbf{x}}$ and this completes the proof. ∎