Achieving Approximate Symmetry Is Exponentially Easier than Exact Symmetry

Behrooz Tahmasebi
Harvard University

Melanie Weber
Harvard University

Abstract

Enforcing *exact symmetry* in machine learning models often yields significant gains in scientific applications, serving as a powerful inductive bias. However, recent work suggests that relying on *approximate symmetry* can offer greater flexibility and robustness. Despite promising empirical evidence, there has been little theoretical understanding, and in particular, a direct comparison between exact and approximate symmetry is missing from the literature. In this paper, we initiate this study by asking: *What is the cost of enforcing exact versus approximate symmetry?* To address this question, we introduce *averaging complexity*, a framework for quantifying the cost of enforcing symmetry via averaging. Our main result is an exponential separation: under standard conditions, achieving exact symmetry requires linear averaging complexity, whereas approximate symmetry can be attained with only logarithmic averaging complexity. To the best of our knowledge, this provides the first theoretical separation of these two cases, formally justifying why approximate symmetry may be preferable in practice. Beyond this, our tools and techniques may be of independent interest for the broader study of symmetries in machine learning.

1 Introduction

The field of *geometric machine learning* aims to incorporate *structures* observed in scientific data into abstract machine learning models, with the goal of leveraging these strong inductive biases to make learning more robust, efficient, and interpretable [4, 47]. Prominent examples include permutation symmetries in point clouds, sign-flip symmetries in spectral graph methods, rotational symmetry in robotics, and structures in molecular and atomistic data [3, 42, 25, 19].

A natural approach to handling symmetries is to encode them *exactly* into the model. The literature offers a variety of such methods, including model-agnostic approaches such as group averaging, data augmentation, canonicalization, and frame averaging [29, 21, 2, 15, 22, 38, 39, 10, 34], as well as model-dependent approaches such as convolutional neural networks and networks with equivariant weights [6, 7, 18, 32, 23, 20, 52].

However, *exact symmetries* come with caveats: in many applications, invariance is only partial and targets respect symmetry only approximately [12, 30, 40, 17, 26, 44]. Another case arises when only partial knowledge of the underlying symmetries is available and symmetry discovery is performed [50, 41, 9, 8, 33, 51, 14]. In this setting, enforcing exact invariance limits universality and expressive power; allowing the model to violate symmetry up to a degree can improve performance while still exploiting strong inductive biases. Motivated by these considerations, researchers have proposed using *approximate symmetry* instead of exact symmetry.

Despite these practical successes, theoretical gaps remain. Since approximate symmetry is a relaxed form of invariance, one might expect it to be *easier* to achieve in data; lower enforcement complexity can yield better sample or computational efficiency and robustness to noise and distributional shifts.

Proceedings of the 1st Conference on Topology, Algebra, and Geometry in Data Science (TAG-DS 2025).

Motivated by these considerations, we study the question: *Is it easier, from a complexity perspective, to enforce approximate symmetry compared to exact symmetry?* We introduce a natural measure for comparing the two regimes: *averaging complexity*.

In averaging complexity, we assume access to a black-box model, and the learner is only allowed to post-process this model linearly through a number of action queries (AQ); the number required by an averaging scheme is its averaging complexity, interpreted as the learner's budget.

> **Main result (informal; under mild conditions).** The averaging complexity of achieving exact symmetry scales linearly with the group size, while approximate symmetry requires only logarithmic complexity in the group size.

This result explains why approximate symmetry is often preferred in practice: in an abstract model, it is *exponentially* easier to achieve. The central message is the exponential separation, indicating that for a given budget, approximate symmetry can achieve stronger results (e.g., in symmetry discovery). Beyond this, our abstract framework and complexity notion, together with the representation-theoretic tools developed here, apply more broadly to geometric machine learning.

2 Related Work

Symmetries appear in many scientific datasets, and equivariant machine learning has proven powerful across applications in particle physics [3], robotics [42], and quantum physics, in both exact [25] and approximate forms [19]. Incorporating symmetry has been shown to improve sample complexity and generalization [43, 36, 11], estimation [5, 37], and learning complexity [16, 35]. Generalization benefits have been observed even when only approximate symmetry holds [28].

Many architectures have been proposed for incorporating symmetries in neural networks, including group-equivariant CNNs [6] and steerable CNNs [7], both built on top of standard convolutional networks [18]. Equivariant graph neural networks [32, 23] and transformers [20] have also been proposed and used in practice. A canonical example for permutation symmetry is Deep Sets [52].

Beyond exact methods, many approaches for introducing relaxed invariance have been proposed in the literature, including modified filters [40], soft equivariance [17, 12], partial equivariance [30], and Lie-algebraic parameterizations [24]. Approximate symmetry has proved effective in reinforcement learning [26] via approximately equivariant Markov decision processes (MDPs). Other examples include the use of structured matrices [31] and relaxed constraints [27]; see also [44, 48]. For neural processes, Ashman et al. [1] propose approximately equivariant schemes with promising benefits. This line of work extends to approximately equivariant graph networks [13] and symmetry breaking for relaxed equivariance [46, 45]. The role and benefits of approximate equivariance in the neural-network optimization landscape have also been studied [49]. In the context of symmetry discovery, many results use semi-supervised methods to learn the underlying symmetry [50, 41, 9, 8, 33, 51, 14].

For model-agnostic methods for equivariant learning, see frame averaging [29, 21, 2] and canonicalization [15, 22, 38, 39, 10, 34] as two widely applicable paradigms.

3 Conclusion and Future Directions

We presented a theoretical study of learning with symmetries, focusing on why *approximate* symmetry is both more convenient in practice and more reasonable for natural data. We introduced an abstract framework that defines the *averaging complexity* of enforcing exact or approximate symmetry as the minimum number of interactions with an oracle via action queries (AQs). Our main result shows an exponential separation: enforcing symmetry exactly can require linear complexity in $|G|$, whereas relaxing to approximate symmetry reduces the complexity to logarithmic in $|G|$, providing theoretical evidence for a sharp gap between the two regimes.

Several directions remain open. First, while this work focuses on finite groups, extending the framework and bounds to infinite groups is both natural and challenging, likely requiring ideas beyond those used here. Second, it would be valuable to leverage our abstract formulation, together with representation-theoretic methods, to analyze other theoretical problems in machine learning under symmetry, such as data augmentation. We leave these questions to future work.

4 Problem Statement

Preliminaries, Notation, and Background. Given $n \in \mathbb{N}$, we write $[n] := \{1, 2, \ldots, n\}$. Let $\mathcal{X}$ be a complete topological space (the data domain), and let G be a finite group. Let $L^2(\mathcal{X})$ denote the space of square-integrable functions on $\mathcal{X}$, assuming $\mathcal{X}$ is equipped with a canonical Borel measure μ.

A (left) *group action* of G on $\mathcal{X}$ is a map $\theta : G \times \mathcal{X} \to \mathcal{X}$ such that $\theta(gh, x) = \theta\big(g, \theta(h, x)\big)$ for all $g, h \in G$ and $x \in \mathcal{X}$, and the identity element of G acts trivially (via the identity map $x \mapsto x$) on $\mathcal{X}$. We write $gx := \theta(g, x)$; for each g, the map $x \mapsto gx$ is a homeomorphism of $\mathcal{X}$. Indeed, without loss of generality, we assume that the canonical measure μ on $\mathcal{X}$ is invariant under the action of G.

Let $\mathcal{F} \subseteq L^2(\mathcal{X})$ be a finite-dimensional real vector space of functions on $\mathcal{X}$, and let $GL(\mathcal{F})$ denote the group of invertible linear mappings from $\mathcal{F}$ to itself (under composition). Assume that for every $g \in G$ and $f \in \mathcal{F}$, the function $x \mapsto f(gx)$ also belongs to $\mathcal{F}$. Define $\rho : G \to GL(\mathcal{F})$ as the canonical group action on $\mathcal{F}$ by leveraging the action on the domain:

$$(\rho(g)[f])(x) := f\left(g^{-1}x\right), \qquad \forall f \in \mathcal{F},\ \forall x \in \mathcal{X}.$$

Indeed, ρ is a (linear) group representation of G on $\mathcal{F}$, meaning that $\rho(gh) = \rho(g)\rho(h)$ under the composition of linear maps.

Averaging Schemes. Consider an abstract setting where a learner aims to post-process the function class $\mathcal{F}$ to enforce a condition (e.g., symmetry). The learner is informed that an arbitrary function $f \in \mathcal{F}$ has been chosen by an oracle and that it remains unchanged throughout post-processing. The learner then issues functional queries to the oracle as follows. Given $f \in \mathcal{F}$ and a group element $g \in G$, the oracle returns the transformed function $x \mapsto f(gx) \in \mathcal{F}$. Because each query evaluates f on $gx \in \mathcal{X}$, we call it an *action query (AQ)*.

After issuing a number of action queries, the learner forms a linear combination of the oracle responses to obtain a post-processed function. The learner has a limited budget and seeks to minimize the number of action queries. This motivates the following definition.

Definition 1 (Averaging Scheme)**.** *An averaging scheme is a function $\omega : G \to \mathbb{R}$ on the finite group G such that $\|\omega\|_{\ell_1(G)} := \sum_{g \in G} \omega(g) = 1$. For a function class $\mathcal{F}$, the averaging operator induced by ω, denoted $\mathbb{E}_\omega : \mathcal{F} \to \mathcal{F}$, is defined by $(\mathbb{E}_\omega[f])(x) := \sum_{g \in G} \omega(g)\, f\left(g^{-1}x\right)$, for all $f \in \mathcal{F}, x \in \mathcal{X}$. The* size *of an averaging scheme is the number of nonzero weights:* $\mathrm{size}(\omega) := \#\{\, g \in G : \omega(g) \neq 0 \,\}$.

Intuitively, an averaging scheme specifies weights used to linearly combine the transformed functions $x \mapsto f\left(g^{-1}x\right)$ to produce the final output. Crucially, averaging schemes *do not* depend on the domain point $x \in \mathcal{X}$; otherwise, they become instances of (weighted) frame averaging, and the notion of averaging complexity becomes ill-defined. We therefore focus on *universal* linear combinations as outputs of averaging operators.

Averaging Complexity. In this paper, we consider the abstract setting where the learner aims to obtain either an exactly symmetric function or an approximately symmetric one. To define *averaging complexity*, we first introduce a few definitions, starting with exact symmetry.

Definition 2 (Exact Symmetry)**.** *A function $f \in \mathcal{F}$ is* exactly symmetric *if, for all $g \in G$ and all $x \in \mathcal{X}$, one has $f(gx) = f(x)$.*

To define approximate symmetry, one must fix a notion of distance from symmetry and allow a relaxation within a prescribed precision. A natural choice is to shrink the "non-symmetry" components of functions (in $L^2(\mathcal{X})$) by a small factor $\epsilon > 0$. When $\epsilon = 0$, the definition reduces to exact symmetry. This formulation also aligns with practice. In this paper, we use two types of approximate symmetry: *weak* and *strong*.

Definition 3 (Weak Approximate Symmetry Enforcement)**.** *An averaging scheme $\omega : G \to \mathbb{R}$* enforces weak approximate symmetry *with respect to a parameter $\epsilon > 0$ if and only if, for every function $f \in \mathcal{F}$, we have*

$$\mathbb{E}_g\left[\int_{\mathcal{X}} |(\mathbb{E}_\omega[f])(x) - (\mathbb{E}_\omega[f])(gx)|^2 \, d\mu(x)\right] \;\leq\; \epsilon\, \mathbb{E}_g\left[\int_{\mathcal{X}} |f(x) - f(gx)|^2 \, d\mu(x)\right],$$

where $g \in G$ is chosen uniformly at random and μ is the canonical Borel measure on $\mathcal{X}$.

Definition 4 (Strong Approximate Symmetry Enforcement)**.** *An averaging scheme* $\omega : G \to \mathbb{R}$ enforces strong approximate symmetry *with respect to a parameter* $\epsilon > 0$ *if and only if, for every function* $f \in \mathcal{F}$*, we have*

$$\int_{\mathcal{X}} |(\mathbb{E}_\omega[f])(x) - (\mathbb{E}_\omega[f])(gx)|^2 \, d\mu(x) \;\leq\; \epsilon \, \mathbb{E}_g\left[\int_{\mathcal{X}} |f(x) - f(gx)|^2 \, d\mu(x)\right], \qquad \forall g \in G,$$

where $g \in G$ *is chosen uniformly at random and* μ *is the canonical Borel measure on* $\mathcal{X}$*.*

In the weak notion, $\mathbb{E}_\omega[f]$ is multiplicatively ϵ-closer (in $L^2(\mathcal{X})$) to being symmetric *on average* over group elements $g \in G$. In the strong notion, the same closeness must hold *for every* $g \in G$.

We are now ready to define the concept of averaging complexity.

Definition 5 (Averaging Complexity)**.** *The* averaging complexity *of enforcing exact, weak approximate, or strong approximate symmetry, denoted* $\mathsf{AC}^{\mathrm{ex}}(\mathcal{F})$*,* $\mathsf{AC}^{\mathrm{wk}}(\mathcal{F}, \varepsilon)$*, and* $\mathsf{AC}^{\mathrm{st}}(\mathcal{F}, \varepsilon)$*, respectively, is the minimal size of an averaging scheme that a learner can construct such that the resulting post-processed function is exactly, weakly approximately, or strongly approximately symmetric, respectively. Formally,*

$$\mathsf{AC}^{\mathrm{ex}}(\mathcal{F}) := \min_{\omega} \Big\{ \mathrm{size}(\omega) : \; (\mathbb{E}_\omega[f])(gx) = (\mathbb{E}_\omega[f])(x) \;\; \forall f \in \mathcal{F},\, g \in G,\, x \in \mathcal{X} \Big\},$$

$$\begin{aligned} \mathsf{AC}^{\mathrm{wk}}(\mathcal{F}, \varepsilon) := \min_{\omega} \Big\{ \mathrm{size}(\omega) : \; & \mathbb{E}_g\big[\|(\mathbb{E}_\omega[f])(x) - (\mathbb{E}_\omega[f])(gx)\|^2_{L^2(\mathcal{X})}\big] \\ & \leq \varepsilon \, \mathbb{E}_g\big[\|f(x) - f(gx)\|^2_{L^2(\mathcal{X})}\big],\; \forall f \in \mathcal{F} \Big\}, \end{aligned}$$

$$\begin{aligned} \mathsf{AC}^{\mathrm{st}}(\mathcal{F}, \varepsilon) := \min_{\omega} \Big\{ \mathrm{size}(\omega) : \; & \|(\mathbb{E}_\omega[f])(x) - (\mathbb{E}_\omega[f])(gx)\|^2_{L^2(\mathcal{X})} \\ & \leq \varepsilon \, \mathbb{E}_g\big[\|f(x) - f(gx)\|^2_{L^2(\mathcal{X})}\big],\; \forall f \in \mathcal{F},\, \forall g \in G \Big\}. \end{aligned}$$

5 Main Results

The main purpose of this paper is to study how various notions of averaging complexity relate to properties of the group action and the function class, and whether there is a fundamental separation between exact and approximate symmetry. Such a separation would show that approximate symmetry is, in an abstract setting, fundamentally easier to achieve.

Assumptions and Definitions. We note that any form of averaging complexity can always be upper bounded *linearly* by the group size via the trivial averaging scheme that queries all group elements $g \in G$. This motivates the question of when *sublinear* averaging complexity is achievable.

To this end, the role of the function class is crucial: trivial classes, such as the set of constant functions, always have trivial averaging complexity. To avoid pathological cases, we assume the following conditions for the domain, group action, and function class:

Assumption 6 (Faithful Group Action)**.** For every nontrivial group element $g \in G$, the map $x \mapsto gx$ is not the identity on the domain.

Assumption 7 (Orbit Separability)**.** For any domain element $x \in \mathcal{X}$ and any nontrivial group element $g \in G$, there exists a function $f \in \mathcal{F}$ such that $f(gx) \neq f(x)$.

These assumptions exclude degenerate examples while remaining sufficiently general. We next define tensor powers of a function class, which we use later in our results.

Definition 8 (Tensor Powers of Function Spaces)**.** *Let* $\mathcal{F}$ *be a finite-dimensional vector space of functions on a domain* $\mathcal{X}$ *and let* $k \in \mathbb{N}$*. Define* $\mathcal{F}^{\otimes k} \;:=\; \mathrm{span}\Big\{ \prod_{i=1}^{k} f_i(x) \;:\; f_i \in \mathcal{F} \text{ for } i \in [k] \Big\}$*, where* $\widetilde{\mathcal{F}^{\otimes k}} \;:=\; \bigoplus_{\ell=0}^{k} \mathcal{F}^{\otimes \ell}$*, and* $\mathcal{F}^{\otimes 0}$ *is the one-dimensional space of constant functions on* $\mathcal{X}$*.*

The construction above uses the base function class $\mathcal{F}$ to form the enlarged class $\widetilde{\mathcal{F}^{\otimes k}}$, which consists of linear combinations of pointwise products of up to k functions from $\mathcal{F}$. In particular, $\mathcal{F}^{\otimes 1} = \mathcal{F}$, and higher orders $k \in \mathbb{N}$ include progressively higher-order polynomial features.

A canonical example is $\mathcal{X} = \mathbb{R}^d$ with $\mathcal{F}$ the set of linear functions on $\mathbb{R}^d$. In this case, $\widetilde{\mathcal{F}^{\otimes k}}$ is exactly the space of polynomials in x of total degree at most k. Another example arises in kernel methods: starting from a base kernel (and its feature map), one may form polynomial feature expansions, which correspond to tensor powers of the base feature space and yield increased expressivity. In this paper, tensor powers serve as a tool for proving lower bounds on the averaging complexity of enforcing exact symmetry. Our goal is to exhibit relatively low degrees k (i.e., low-order polynomial features) for which the required averaging complexity is linear in $|G|$.

Theorem 9 (Averaging Complexity of Exact Symmetry Enforcement)**.** *Under the above assumptions, for any function class $\mathcal{F}$ there exists an integer K such that the averaging complexity of exact symmetry enforcement is $\mathsf{AC}^{\mathrm{ex}}(\widetilde{\mathcal{F}^{\otimes k}}) = |G|$ for all $k \geq \mathsf{K}$.*

Theorem 9 asserts that exact symmetry requires *linear* averaging complexity once polynomial features of degree $k = \mathsf{K}$ are included. Even though the bound $\mathsf{K} \leq |G|$ holds trivially, one can often obtain much sharper bounds.

Theorem 10 (Averaging Complexity of Approximate Symmetry Enforcement)**.** *For any function class $\mathcal{F}$ and any $\varepsilon > 0$, the averaging complexities of weak and strong approximate symmetry enforcement satisfy* $\mathsf{AC}^{\mathrm{st}}(\mathcal{F}, \varepsilon) = \mathcal{O}\left(\frac{\log|G|}{\varepsilon}\right)$, $\mathsf{AC}^{\mathrm{wk}}(\mathcal{F}, \varepsilon) = \mathcal{O}\left(\frac{\log|G|}{\varepsilon}\right)$.

These bounds hold *uniformly* for all function classes, without Assumptions 6 and 7, whether or not tensor powers are used. Thus, the upper bounds for approximate symmetry enforcement are *universal*. In particular, they apply to the classes considered in Theorem 9, for which exact symmetry requires linear averaging complexity. Therefore, approximate symmetry enforcement needs only *logarithmic* averaging complexity (in $|G|$), yielding an *exponential* separation between the approximate and exact regimes.

For proofs, please refer to the full version of this paper.

Acknowledgments

BT and MW were partially supported by NSF Award CBET-2112085. MW acknowledges partial funding from NSF award DMS-2406905 and a Sloan Research Fellowship.

References

[1] Matthew Ashman, Cristiana Diaconu, Adrian Weller, Wessel Bruinsma, and Richard Turner. Approximately equivariant neural processes. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2

[2] Matan Atzmon, Koki Nagano, Sanja Fidler, Sameh Khamis, and Yaron Lipman. Frame averaging for equivariant shape space learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2

[3] Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi, David Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics. In *Int. Conference on Machine Learning (ICML)*, 2020. 1, 2

[4] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021. 1

[5] Ziyu Chen, Markos Katsoulakis, Luc Rey-Bellet, and Wei Zhu. Sample complexity of probability divergences under group symmetry. In *Int. Conference on Machine Learning (ICML)*, 2023. 2

[6] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Int. Conference on Machine Learning (ICML)*, 2016. 1, 2

[7] Taco S Cohen and Max Welling. Steerable cnns. In *Int. Conference on Learning Representations (ICLR)*, 2017. 1, 2

[8] Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1, 2

[9] Krish Desai, Benjamin Nachman, and Jesse Thaler. Symmetry discovery with deep learning. *Physical Review D*, 105(9):096031, 2022. 1, 2

[10] Nadav Dym, Hannah Lawrence, and Jonathan W Siegel. Equivariant frames and the impossibility of continuous canonicalization, 2024. 1, 2

[11] Bryn Elesedy. Provably strict generalisation benefit for invariance in kernel methods. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 2

[12] Marc Finzi, Gregory Benton, and Andrew G Wilson. Residual pathway priors for soft equivariance constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1, 2

[13] Ningyuan Huang, Ron Levie, and Soledad Villar. Approximately equivariant graph networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 2

[14] Dongsung Huh. Discovering group structures via unitary representation learning. In *Int. Conference on Learning Representations (ICLR)*, 2025. 1, 2

[15] Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *Int. Conference on Machine Learning (ICML)*, 2023. 1, 2

[16] Bobak T Kiani, Thien Le, Hannah Lawrence, Stefanie Jegelka, and Melanie Weber. On the hardness of learning under symmetries. In *Int. Conference on Learning Representations (ICLR)*, 2024. 2

[17] Hyunsu Kim, Hyungi Lee, Hongseok Yang, and Juho Lee. Regularizing towards soft equivariance under mixed symmetries. In *Int. Conference on Machine Learning (ICML)*, 2023. 1, 2

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2012. 1, 2

[19] Dominik S Kufel, Jack Kemp, DinhDuy Vu, Simon M Linsel, Chris R Laumann, and Norman Y Yao. Approximately symmetric neural networks for quantum spin liquids. *Physical Review Letters*, 135(5):056702, 2025. 1, 2

[20] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *Int. Conference on Learning Representations (ICLR)*, 2023. 1, 2

[21] Yuchao Lin, Jacob Helwig, Shurui Gui, and Shuiwang Ji. Equivariance via minimal frame averaging for more symmetries and efficiency. In *Int. Conference on Machine Learning (ICML)*, 2024. 1, 2

[22] George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 1, 2

[23] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *Int. Conference on Learning Representations (ICLR)*, 2019. 1, 2

[24] Daniel McNeela. Almost equivariance via lie algebra convolutions. In *Symmetry and Geometry in Neural Representations (NeurReps) Workshop*, 2023. 2

[25] Quynh T Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J Coles, Frederic Sauvage, Martin Larocca, and Marco Cerezo. Theory for equivariant quantum neural networks. *PRX Quantum*, 5(2):020328, 2024. 1, 2

[26] Jung Yeon Park, Sujay Bhatt, Sihan Zeng, Lawson L.S. Wong, Alec Koppel, Sumitra Ganesh, and Robin Walters. Approximate equivariance in reinforcement learning. In *Int. Conference on Artificial Intelligence and Statistics (AISTATS)*, 2025. 1, 2

[27] Stefanos Pertigkiozoglou, Evangelos Chatzipantazis, Shubhendu Trivedi, and Kostas Daniilidis. Improving equivariant model training via constraint relaxation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2

[28] Mircea Petrache and Shubhendu Trivedi. Approximation-generalization trade-offs under (approximate) group equivariance. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 2

[29] Omri Puny, Matan Atzmon, Edward J Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *Int. Conference on Learning Representations (ICLR)*, 2022. 1, 2

[30] David W Romero and Suhas Lohit. Learning partial equivariances from data. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1, 2

[31] Ashwin Samudre, Mircea Petrache, Brian Nord, and Shubhendu Trivedi. Symmetry-based structured matrices for efficient approximately equivariant networks. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. 2

[32] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Int. Conference on Machine Learning (ICML)*, 2021. 1, 2

[33] Ben Shaw, Abram Magner, and Kevin Moon. Symmetry discovery beyond affine transformations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 1, 2

[34] Zakhar Shumaylov, Peter Zaika, James Rowbottom, Ferdia Sherry, Melanie Weber, and Carola-Bibiane Schönlieb. Lie algebra canonicalization: Equivariant neural operators under arbitrary lie groups. In *Int. Conference on Learning Representations (ICLR)*, 2025. 1, 2

[35] Ashkan Soleymani, Behrooz Tahmasebi, Stefanie Jegelka, and Patrick Jaillet. Learning with exact invariances in polynomial time. In *Int. Conference on Machine Learning (ICML)*, 2025. 2

[36] Behrooz Tahmasebi and Stefanie Jegelka. The exact sample complexity gain from invariances for kernel regression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 2

[37] Behrooz Tahmasebi and Stefanie Jegelka. Sample complexity bounds for estimating probability divergences under invariances. In *Int. Conference on Machine Learning (ICML)*, 2024. 2

[38] Behrooz Tahmasebi and Stefanie Jegelka. Regularity in canonicalized models: A theoretical perspective. In *Int. Conference on Artificial Intelligence and Statistics (AISTATS)*, 2025. 1, 2

[39] Behrooz Tahmasebi and Stefanie Jegelka. Generalization bounds for canonicalization: A comparative study with group averaging. In *Int. Conference on Learning Representations (ICLR)*, 2025. 1, 2

[40] Tycho van der Ouderaa, David W Romero, and Mark van der Wilk. Relaxing equivariance constraints with non-stationary continuous filters. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1, 2

[41] Tycho van der Ouderaa, Alexander Immer, and Mark van der Wilk. Learning layer-wise equivariances automatically using gradients. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 1, 2

[42] Dian Wang, Robin Walters, and Robert Platt. $SO(2)$-equivariant reinforcement learning. In *Int. Conference on Learning Representations (ICLR)*, 2022. 1, 2

[43] Rui Wang, Robin Walters, and Rose Yu. Incorporating symmetry into deep dynamics models for improved generalization. In *Int. Conference on Learning Representations (ICLR)*, 2021. 2

[44] Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *Int. Conference on Machine Learning (ICML)*, 2022. 1, 2

[45] Rui Wang, Robin Walters, and Tess Smidt. Relaxed octahedral group convolution for learning symmetry breaking in 3d physical systems. In *NeurIPS AI for Science Workshop*, 2023. 2

[46] Rui Wang, Elyssa Hofgard, Han Gao, Robin Walters, and Tess Smidt. Discovering symmetry breaking in physical systems with relaxed group convolution. In *Int. Conference on Machine Learning (ICML)*, 2024. 2

[47] Melanie Weber. Geometric machine learning. *Wiley Online Library*, 2025. 1

[48] Zhiqiang Wu, Yingjie Liu, Licheng Sun, Jian Yang, Hanlin Dong, Shing-Ho J Lin, Xuan Tang, Jinpeng Mi, Bo Jin, and Xian Wei. Relaxed rotational equivariance via g-biases in vision. In *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, 2025. 2

[49] YuQing Xie and Tess Smidt. A tale of two symmetries: Exploring the loss landscape of equivariant models. *arXiv preprint arXiv:2506.02269*, 2025. 2

[50] Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In *Int. Conference on Machine Learning (ICML)*, 2023. 1, 2

[51] Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. Latent space symmetry discovery. In *Int. Conference on Machine Learning (ICML)*, 2024. 1, 2

[52] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 1, 2