What Makes a Machine Learning Task a Good Candidate for an Equivariant Network?

Scott Mahan [1] **Davis Brown** [1] **Timothy Doster** [1] **Henry Kvinge** [1 2]

Abstract

Because of the prevalence of symmetry in real-world data, the development of deep learning architectures that incorporate this structure into network design has become an important area of research. Empirically however, equivariant architectures tend to provide more benefit in some settings than others. Since the development of new equivariant layers is a substantial research task and existing equivariant architectures tend to be more complex and harder for the non-expert to work with, identifying those situations where architectural equivariance is likely to bring the most benefit is an important question for the practitioner. In this short paper we begin to explore this question. Our preliminary studies suggest that (i) equivariant architectures are more useful when groups are more complex and data is more high-dimensional, (ii) aligning the type of equivariance with the symmetries in the task brings the most benefit, (iii) equivariant architectures tend to be beneficial across data regimes, and (iv) equivariant architectures display similar scaling behavior (as a function of training set size) as non-equivariant architectures.

1. Introduction

The symmetries in data can be used to impose useful inductive biases on deep learning architectures. This is often done by developing layers that are equivariant to the relevant symmetry group while at the same time remaining as expressive as possible. The convolutional neural network architecture, which continues to be central in computer vision, is the most famous example of this, encoding translation equivariance in images and related modalities. There are many other symmetries in data however, some of which are substantially more complex than this example (e.g., even moving from the orthogonal group $O(2)$ for 2-dimensional images, to $O(3)$ in 3-dimensional data is a significant jump in mathematical complexity). As equivariant architectures tend to be more complicated to work with than more common architectures and developing a new equivariant architecture is often a major research effort, understanding if and when one should consider using an equivariant architecture is useful. In this paper we pose the question:

What characteristics of a machine learning task imply that an equivariant architecture will perform better than a non-equivariant architecture?

Our initial motivation came from the anecdotal observation that while rotation equivariant architectures bring only marginal gains to computer vision tasks involving 2d images (Kvinge et al., 2022) (at least in the large training set regime), they appear to be essential to tasks involving 3d data (e.g., molecular modeling) (Duval et al., 2023). This makes intuitive sense; more complicated symmetries would benefit more from built in priors. But since the role of data symmetry in deep learning models is still not fully understood, it is interesting to create experiments to explore whether this trend holds in other settings where the dimension/group size associated with a task can be varied beyond $n = 2$ and 3. To begin to disentangle the factors impacting equivariant architecture performance, we identify several elementary properties of a machine learning task that we conjecture to be important. These include (i) the dimension of the input/complexity of the group, (ii) the extent to which the symmetries captured by the architecture align with the symmetries of the task, and (iii) the volume of training data. Properties to explore in a future version of this work include Lie group dimension and dataset diversity.

In our first experiment we look at the problem of predicting N-body dynamics using a standard graph neural network (GNN) and a $E(n)$-equivariant GNN where n is the dimension in which the particles move and $E(n)$ is the Eu-

[1]Pacific Northwest National Laboratory [2]University of Washington. Correspondence to: Scott Mahan <scott.mahan@pnnl.gov>.

Accepted as an extended abstract for the Geometry-grounded Representation Learning and Generative Modeling Workshop at the 41st International Conference on Machine Learning, ICML 2024, Vienna, Austria.

clidean group for $\mathbb{R}^n$. Varying n, we see that as the input dimension/group size increases the benefits of using an equivariant architecture increase relative to a non-$E(n)$-equivariant architecture. This aligns with the observation about 2d-images and 3d-data. In our second experiment we train a range of equivariant convolutional neural networks (CNNs) on rotated MNIST and flipped and rotated MNIST. Each model is composed of layers equivariant to a different subgroup of the orthogonal group $O(2)$ (e.g., cyclic groups C_r, dihedral groups D_r, $SO(2)$, etc.). We also vary the available training data. We find, as might be expected, that equivariant models do best when the group they are equivariant to aligns with the symmetries in the task. More interestingly, we find that equivariant architectures provide a benefit across training set size regimes, contradicting folk wisdom that built-in equivariance is only important when there are few examples of each class available. Finally, we observe that the scaling plots of equivariant and non-equivariant architectures have similar slopes as a function of training data, suggesting that when trained on the same data and when models are the same approximate size, equivariant and non-equivariant models tend to have similar scaling properties.

Finally, it is important to note that this paper represents work in progress. As such, most lines of inquiry will need further experiments and analysis to establish firm conclusions (especially when it comes to getting coverage of a broader range of groups and applications). We hope that the ideas presented here will serve as impetus for further research into this question and related questions.

Contributions: In summary, this short paper contributes the following: (1) We pose the problem of identifying when an ML task will benefit from the use of equivariant architectures, focusing on easily measurable problem parameters. (2) We run experiments to explore how the dimension of the input data, the order of the group (if finite), and the amount of data available impact the performance of an equivariant model vs. a non-equivariant model. (3) We summarize some initial trends that we observe. For instance, we see the largest gains from equivariance with higher dimensional data and more complex symmetry groups and notice that equivariant models outstrip their non-equivariant counterparts in both small- and large-data regimes. Finally, we find evidence that the slope of the scaling curves for equivariant and non-equivariant models are very similar.

2. Equivariance and ML Tasks

Let G be a group with real representations V and W, so that V and W are $\mathbb{R}$-vector spaces and there are group homomorphisms $\phi_1 : G \to GL(V)$ and $\phi_2 : G \to GL(W)$. When it is understood from the context we will omit ϕ_1 and ϕ_2 and just write $gv := \phi_1(g)v$ (and analogously for ϕ_2 and W). If $f : V \to W$ is a neural network then we say that f is G-equivariant if for any $g \in G$ and any $v \in V$, $f(gv) = gf(v)$[1]. Since f is a neural network, it has a decomposition into l layers which we write as $f = f_l \circ \cdots \circ f_1$ with $f_i : V_{i-1} \to V_i$.

One of several approaches to building equivariance into a network is to construct the network from equivariant layers. Abusing terminology, we will refer to this type of network as an equivariant architecture. By now, G-equivariant layers have been designed for a range of groups G and representations of G. In this work our experiments focus on some more common examples including: (i) the cyclic groups C_r which are generated by a rotation of $360°/r$ degrees in the plane, (ii) the dihedral groups D_r which are generated by a rotation of $360°/r$ degrees and a reflection across the x-axis in the plane, (iii) $O(2)$, which is the Lie group of all rotations and reflections in the plane (note that C_r, D_r for all $r \geq 1$ are subgroups of $O(2)$), and (iv) $E(n)$, which is generated by all translations, reflections, and rotations of $\mathbb{R}^n$.

3. Related Work

The concepts of invariance and equivariance which permeate mathematics and physics are becoming increasingly central to machine learning. This has been driven by the expanding number of applications that deep learning methods are used in, particularly those coming from the sciences. There is now a diverse suite of approaches for constructing invariant or equivariant networks. Some methods involve manipulation of the input data by either transforming it into a canonical example (Kaba et al., 2023) or simply through targeted data augmentation (Hauberg et al., 2016; Cubuk et al., 2018). In this work we focus on equivariant architectures (Cohen & Welling, 2016a; Kondor & Trivedi, 2018; Cohen & Welling, 2016b). They enforce equivariance by composing a network from equivariant layers. There are now equivariant architectures for a broad range of groups and representations. These range from cyclic groups C_r, dihedral groups D_{2r}, and $SO(2)$ applied to 2d images (Weiler & Cesa, 2019) to $SO(3)$ applied to 3d point clouds and volumetric data (Weiler et al., 2018) or molecular models (Duval et al., 2023) to the symmetric groups applied to elements of a set (Zhang et al., 2019) or vertices of a graph (Scarselli et al., 2009).

There is also a rich literature characterizing the theoretical and empirical properties of equivariance in deep learning models. Such studies include (Gruver et al., 2022) which analyzes the extent to which equivariance is lost at specific

[1]Note that invariance is a special case of equivariance where the action of G on W is trivial.

layers and (Kvinge et al., 2022) which attempts to quantify the extent to which the internal representations and final output of a model are equivariant. Other work focuses on the theoretical properties of invariance in deep learning models (Bloem-Reddy et al., 2020; Chen et al., 2020; Singla et al., 2021) and the universality of invariant neural networks (Maron et al., 2019; Yarotsky, 2022). However, despite the widespread belief that equivariant models improve generalization on equivariant data distributions, there are few results demonstrating this exact result.

Among such results, (Elesedy & Zaidi, 2021) provides provably strict generalization benefits from equivariant models when the target distribution is equivariant *and linear* (i.e., $X \sim \mathcal{N}(0, \sigma^2 I)$ and $Y = \Theta^T X + \epsilon$). They show that the expected test generalization gap between the least-squares estimate Θ_n on n i.i.d. examples and its equivariant counterpart $\overline{\Theta}_n := \int_G \phi_1(g)\Theta_n\phi_2(g^{-1})dg$ is

$$\sigma_\epsilon^2 \frac{\dim(V)\dim(W) - \langle \chi_{\phi_1}, \chi_{\phi_2} \rangle}{n - \dim(V) - 1} > 0, \qquad (1)$$

where $\langle \chi_{\phi_1}, \chi_{\phi_2} \rangle$ is the scalar product between the characters of the representations ϕ_1 and ϕ_2. The expression in (1) provides some insight on factors that affect the generalization benefit from equivariance. Note that $\dim(V)\dim(W)$ is the dimension of the space of linear maps $V \to W$, while $\langle \chi_{\phi_1}, \chi_{\phi_2} \rangle$ is the dimension of the space of equivariant linear maps. (Elesedy & Zaidi, 2021) shows that the numerator in (1) thus represents the dimension of the space of linear maps which vanish when averaged over G, which is a heuristic for the importance of symmetry in the task. In other words, the importance of symmetry directly controls the generalization benefit from equivariance as is believed to be the case in general.

Other results on generalization of equivariant models bound the worst-case generalization error using complexity measures like VC dimension or Rademacher complexity (Sokolic et al., 2017; Sannai et al., 2021). These measures are data-agnostic and only capture the complexity of a class of models. Additionally, (Lyle et al., 2020) provides a PAC-Bayes generalization error bound for equivariant neural networks. Across these works, there is no implication that equivariant models generalize strictly better on equivariant distributions. Only (Elesedy & Zaidi, 2021) shows this result in the linear case. As we present our empirical results, we will compare them to these existing theoretical results on the generalization of equivariant architectures.

4. Experiments

We conduct two experiments to explore the performance of equivariant and non-equivariant models relative to the characteristics of a machine learning task. In the first, we vary the input dimension and associated group "complexity" by training $E(n)$-equivariant and non-equivariant graph neural networks (GNNs) on a dynamical system task for varying values of n. In the second, we look at the performance of a range of rotation equivariant CNNs on two versions of MNIST (Deng, 2012); one where digits have been randomly rotated and one where digits have been randomly flipped and rotated. In this latter experiment we vary both the group and the amount of training data.

4.1. Equivariant architectures and data dimension in N-body dynamical systems

We apply graph neural networks to model N-body dynamical systems in $\mathbb{R}^n$ using the experimental set-up from (Kipf et al., 2018). Each node in the GNN represents a charged particle with n-dimensional position and velocity data, while edges between nodes indicate whether they have the same or opposite charge. Though the N-body problem is traditionally considered in dimensions $n = 2$ or 3, one can set n equal to any positive integer. Such dynamical systems have been modeled with graph neural networks in the context of deep reinforcement learning (Chua et al., 2018). The system we consider is governed by forces between particles, which determine the ending positional configuration after a fixed amount of time. The goal of the machine learning task is to predict the ending configuration from the initial configuration. This system is equivariant with respect to the Euclidean group $E(n)$, which is generated by all translations, rotations, and reflections. That is, rotating, reflecting, or translating the positions of all particles results in the same solution up to an equivalent rotation, reflection, or translation. Thus it makes sense to try to use an $E(n)$-equivariant neural network for this task[2].

For each dimension $n = 2, 3, 4, 6$, and 8, we sample $10,000$ trajectories for training and $2,000$ for testing. The number of particles is fixed at $N = 5$. Each trajectory consists of an input/output graph pair sampled as follows: we randomly initialize positions $\mathbf{x}^{(0)} = \{\mathbf{x}_1^{(0)}, \ldots, \mathbf{x}_5^{(0)}\} \in \mathbb{R}^{5\times d}$, velocities $\mathbf{v}^{(0)} = \{\mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_5^{(0)}\} \in \mathbb{R}^{5\times d}$, and charges $\{c_1, \ldots, c_5\} \in \{-1, 1\}$; next, we apply $T = 10,000$ steps of Euler's method, where change in position is determined by velocity and change in velocity is determined by forces proportional to the relative charge between particles and inversely proportional to the $(n-1)^{\text{th}}$ power of the distance between particles. The output graph's node data consists of the final positions $\mathbf{x}^{(T)}$. Models are trained to minimize the Mean Squared Error between predicted and ground truth positions.

We use a graph neural network as our baseline model (Scarselli et al., 2009). This model does not have $E(n)$-

[2]Note that this system is also equivariant to permutation of particles since all particles are functionally identical, but this equivariance is in all the GNNs that we use so we ignore it.

equivariance for node data (position and velocity), though as noted above it does have permutation equivariance. In particular, we follow the neural message passing framework (Gilmer et al., 2017). The $E(n)$-equivariant version of the model uses $E(n)$-equivariant graph convolutional layers (Satorras et al., 2021), which produce node features *and* positions at the next layer.

Results: To measure the importance of equivariance, we track the best test mean squared error (MSE) achieved by each model in dimension $d = 2, 3, 4, 6, 8$ with $n = 10,000$ training samples. MSE_{GNN} represents the ordinary graph neural network while MSE_{EGNN} represents the equivariant model. Figure 1 plots the ratio $\text{MSE}_{GNN}/\text{MSE}_{EGNN}$ as a function of dimension. We note that $\text{MSE}_{GNN}/\text{MSE}_{EGNN}$ is not only greater than 1, it tends to grow as a function of n indicating (in this setting at least), equivariance does tend to be more valuable for data of higher dimension and/or with more complicated symmetries. We note that since both the dimension of the representation $\mathbb{R}^n$ and the complexity of the group $E(n)$ grow with n, it is unclear if one of these drives the gains seen by the equivariant architecture.

One can ask why we use the quotient of MSE_{GNN} and MSE_{EGNN} rather than, for example, $\text{MSE}_{GNN} - \text{MSE}_{EGNN}$. This choice was made based on the observation that the nature and difficulty of the problem changes as n changes. This is a persistent challenge encountered when running experiments where task parameters such as input data dimension vary. It turns out that for this N-body problem, increasing the dimension changes the underlying dynamics making the task easier in relative terms. Thus, the absolute difference between MSE values decreases, but only because the MSE for both models decreases as n increases. The dependence on n may also explain the interesting spike in Figure 1 at $n = 3$. By plotting the quotient we are able to capture the difference in relative performance between the $E(n)$-equivariant and non-equivariant models.

Takeaway: Equivariant architectures may be more valuable as the input dimension of the data and the complexity of the group of symmetries grow larger.

4.2. Data Scaling and Equivariance Type for Rotated MNIST

For our second experiment, we explore how data volume and the choice of group impact the difference in performance between equivariant and non-equivariant CNNs on rotated MNIST and flipped and rotated MNIST (the latter consisting of compositions of random rotations and a possible reflection). We limit each one to a maximum of $50,000$ training examples and $10,000$ test examples to keep the sizes consistent. All images in rotated MNIST are rotated by a fixed random angle, making it a rotation-invariant task. All images in flipped and rotated MNIST are also randomly flipped making it an $O(2)$-invariant task.

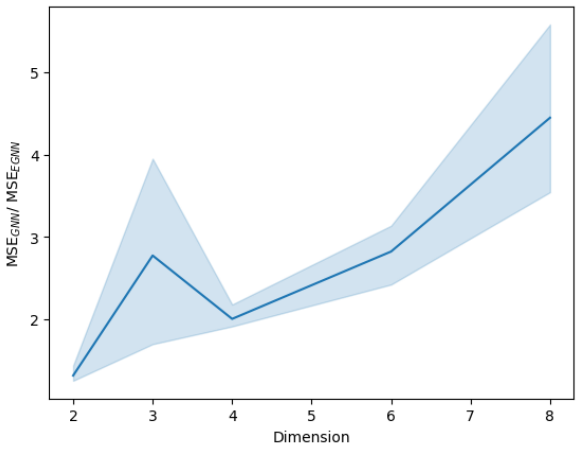

Figure 1. The ratio of the MSE of a model that is not equivariant to the group $E(n)$ over the MSE of an $E(n)$-equivariant model as a function of the input dimension n. Shaded regions indicate 95% confidence values.

We compare ordinary CNNs to group equivariant CNNs, using the $E(n)$-equivariant steerable CNN library (Cesa et al., 2022). We have models that are equivariant to various subgroups of $E(2)$, including C_4, D_4, C_8, D_8, $SO(2)$, and $O(2)$. We will identify these models by their subgroup of $O(2)$, which is $G = \{e\}$ for the conventional CNN. Each convolutional network consists of six convolution blocks, which contains a (possibly equivariant) convolution layer, batch normalization, and ELU activation. We apply average pooling after every two blocks. We have attempted to keep the number of learnable parameters between each model approximately equal. Everything is trained using the Adam optimizer with a learning rate of 5×10^{-5} and a batch size of 64. To assess data scaling, we train each model to completion on $n = 1,000,\ 2,000,\ \ldots,\ 10,000,\ 20,000,\ \ldots,\ 50,000$ training examples.

Results: Figure 2 plots best test error rate vs. the number of training examples for each model on rotated MNIST (top) and flipped and rotated MNIST (bottom). In both cases, we see that the models that actually capture the symmetry of the dataset (exclusively continuous rotational symmetry for rotated MNIST and continuous rotational and reflection symmetry for flipped and rotated MNIST) tend to have lower error in all data regimes. That is, the $SO(2)$-equivariant architecture has the lowest error for rotated MNIST and the $O(2)$-equivariant architecture has the lowest error for flipped and rotated MNIST. Beyond that, we find that the finite groups (cyclic and dihedral) with the highest order which best approximate the underlying continuous group, do better than lower order groups with

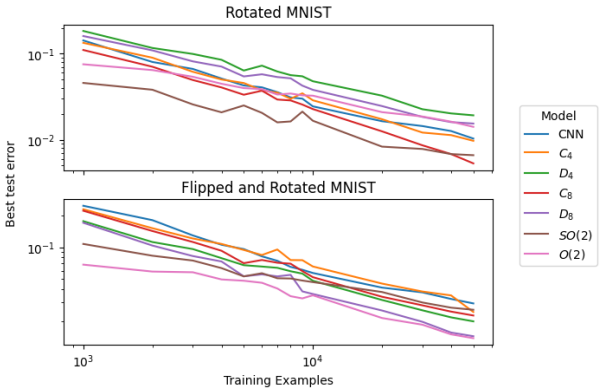

Figure 2. Best test error (misclassification rate) vs. number of training examples for models of various equivariance levels on rotated MNIST and flipped and rotated MNIST.

worse approximation (e.g., C_4 and D_4). Note that this is in contrast to (Weiler & Cesa, 2019), which found that networks equivariant to finite subgroups sometimes perform better than networks equivariant to their continuous parent even if the underlying symmetry is continuous.

Folk wisdom has often suggested that equivariance is most important in the low-data regime. This is natural to expect. If a dataset is small, a model may not see all possible transformations of a given class and built-in equivariance might be more useful. We find that while equivariance is valuable when the training set is small, it provides benefit in all data regimes. This implies that even for problems where ample training data is available, one should not rule out the use of equivariant architectures.

We notice that in Figure 2, each curve is approximately linear (on a log-log scale), agreeing with past empirical work on data scaling laws (Hoffmann et al., 2022). Interestingly, the slopes between models with different types of equivariance are all relatively similar suggesting that changing the G to which a model is equivariant only changes scaling behavior up to a shift in the log-log plot. This is in contrast to behavior seen with models with different numbers of parameters, where the scaling curve slopes can vary. One interesting implication to this is that if we understand the scaling behavior of a non-equivariant model, we only need to calculate one constant term to be able to calculate scaling behavior for any given equivariant model.

Takeaway: Models with equivariant architectures whose symmetries best align with the symmetries in the data tend to perform better for all amounts of training data explored. The scaling properties of equivariant and non-equivariant models tend to be similar up to a shift in the log-log plot.

5. Conclusion

In this work we ask whether it is possible to predict when a given ML task will benefit from the use of an equivariant neural network architecture. We focus on how well the group aligns with symmetries in the data, the group complexity and input dimension, and quantity of training data. Our conclusions suggest that (i) tasks with high-dimensional data and complex symmetries may benefit from equivariant architectures more than low-dimensional data with simple symmetries, (ii) the group that a model is built to be equivariant to should align with the symmetries of the data, and (iii) the scaling properties of equivariant and non-equivariant models are the same for fixed data up to a possible shift in the log-log plot.

We note there are several limitations in this preliminary study. Firstly, too few groups and types of ML tasks are explored. In particular, all experiments in this paper concern a special type of equivariance: invariance. Secondly, strong scaling experiments require extensive hyperparameter searches and carefully controlled training runs. We plan to address these issues in another iteration of this work.

References

Bloem-Reddy, B., Whye, Y., et al. Probabilistic symmetries and invariant neural networks. *Journal of Machine Learning Research*, 21(90):1–61, 2020. 3

Cesa, G., Lang, L., and Weiler, M. A program to build e(n)-equivariant steerable CNNs. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=WE4qe9xlnQw. 4

Chen, S., Dobriban, E., and Lee, J. H. A group-theoretic framework for data augmentation. *Journal of Machine Learning Research*, 21(245):1–71, 2020. 3

Chua, K., Calandra, R., McAllister, R., and Levine, S. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Advances in Neural Information Processing Systems*, NIPS'18, pp. 4759–4770, Red Hook, NY, USA, 2018. Curran Associates Inc. 3

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016a. 2

Cohen, T. S. and Welling, M. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016b. 2

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018. 2

Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6):141–142, 2012. 3

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Liò, P., Bengio, Y., and Bronstein, M. A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511*, 2023. 1, 2

Elesedy, B. and Zaidi, S. Provably strict generalisation benefit for equivariant models. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2959–2969. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/elesedy21a.html. 3

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html. 4

Gruver, N., Finzi, M., Goldblum, M., and Wilson, A. G. The lie derivative for measuring learned equivariance. *arXiv preprint arXiv:2210.02984*, 2022. 2

Hauberg, S., Freifeld, O., Larsen, A. B. L., Fisher, J., and Hansen, L. Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In *Artificial intelligence and statistics*, pp. 342–350. PMLR, 2016. 2

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022. 5

Kaba, S.-O., Mondal, A. K., Zhang, Y., Bengio, Y., and Ravanbakhsh, S. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pp. 15546–15566. PMLR, 2023. 2

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106. 7

Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., and Zemel, R. Neural relational inference for interacting systems. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2688–2697. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/kipf18a.html. 3

Kondor, R. and Trivedi, S. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, pp. 2747–2755. PMLR, 2018. 2

Kvinge, H., Emerson, T., Jorgenson, G., Vasquez, S., Doster, T., and Lew, J. In what ways are deep neural networks invariant and how should we measure this? *Advances in Neural Information Processing Systems*, 35:32816–32829, 2022. 1, 3

Lyle, C., van der Wilk, M., Kwiatkowska, M., Gal, Y., and Bloem-Reddy, B. On the benefits of invariance in neural networks. *arXiv preprint arXiv:2005.00178*, 2020. 3

Maron, H., Fetaya, E., Segol, N., and Lipman, Y. On the universality of invariant networks. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4363–4371. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/maron19a.html. 3

Sannai, A., Imaizumi, M., and Kawano, M. Improved generalization bounds of group invariant / equivariant deep networks via quotient feature spaces. In de Campos, C. and Maathuis, M. H. (eds.), *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 771–780. PMLR, 27–30 Jul 2021. URL https://proceedings.mlr.press/v161/sannai21a.html. 3

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. *CoRR*, abs/2102.09844, 2021. URL https://arxiv.org/abs/2102.09844. 4, 7

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. doi: 10.1109/TNN.2008.2005605. 2, 3

Singla, V., Ge, S., Ronen, B., and Jacobs, D. Shift invariance can reduce adversarial robustness. *Advances in Neural Information Processing Systems*, 34:1858–1871, 2021. 3

Sokolic, J., Giryes, R., Sapiro, G., and Rodrigues, M. Generalization Error of Invariant Classifiers. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1094–1103. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/sokolic17a.html. 3

Weiler, M. and Cesa, G. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019. 2, 5

Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. S. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018. 2

Yarotsky, D. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1):407–474, February 2022. 3

Zhang, Y., Hare, J., and Prugel-Bennett, A. Deep set prediction networks. *Advances in Neural Information Processing Systems*, 32, 2019. 2

A. Experimental details

Equivariance in N-body Dynamical Systems: In our N-body experiments, each model takes initial positions $\mathbf{x}^0 = \mathbf{x}^{(0)}$ as input and $\mathbf{h}_i^0 = \|\mathbf{v}_i^{(0)}\|$ as features. Edge values represent relative charges $e_{ij} = c_i c_j$. We apply 4 convolutional layers in each model and train for 10,000 epochs using the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 5×10^{-4} and a batch size of 128.

Our $E(n)$-equivariant network has convolutions that are defined by

$$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, \|\mathbf{x}_i^l - \mathbf{x}_j^l\|, e_{ij}), \tag{2}$$

$$\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + C \sum_{j \neq i} (\mathbf{x}_j^l - \mathbf{x}_i^l) \phi_x(\mathbf{m}_{ij}), \tag{3}$$

$$\mathbf{m}_i = \sum_{j \neq i} \mathbf{m}_{ij}, \tag{4}$$

$$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i). \tag{5}$$

Note that equation (2) is a function of the distance between nodes, and equation (3) models the influence on node v_i's position from the rest of the nodes. By only relying on the relative position between points, the convolutions in this model are equivariant to rotations, translations, and reflections in $\mathbb{R}^n$. The equivariant graph convolution only requires a small modification to include velocity data and retain equivariance (Satorras et al., 2021).

Our GNN that is not $E(n)$-equivariant is given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $v_i \in \mathcal{V}$ and edges $e_{ij} \in \mathcal{E}$. The graph convolution at layer l is defined by

$$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i, \mathbf{h}_j, e_{ij}), \tag{6}$$

$$\mathbf{m}_i = \sum_{j \neq i} \mathbf{m}_{ij}, \tag{7}$$

$$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i). \tag{8}$$

Here, $\mathbf{h}_i^l \in \mathbb{R}^f$ denotes the feature embedding of node v_i at hidden layer l. The functions ϕ_e and ϕ_h are approximated by Multilayer Perceptrons (MLPs).