Quantifying Lie Group Learning with Local Symmetry Error

Vasco Portilheiro
VASCO.PORTILHEIRO.23@UCL.AC.UK
Gatsby Computational Neuroscience Unit, UCL

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

Abstract

Despite increasing interest in using machine learning to discover symmetries, no quantitative measure has been proposed in order to compare the performance of different algorithms. Our proposal, both intuitively and theoretically grounded, is to compare Lie groups using a *local symmetry error*, based on the difference between their infinitesimal actions at any possible datapoint. Namely, we use a well-studied metric to compare the induced tangent spaces. We obtain an upper bound on this metric which is uniform across datapoints, under some conditions. We show that when one of the groups is a circle group, this bound is furthermore both tight and easily computable, thus globally characterizing the local errors. We demonstrate our proposal by quantitatively evaluating an existing algorithm. We note that our proposed metric has deficiencies in comparing tangent spaces of different dimensions, as well as distinct groups whose local actions are similar.

1. Introduction

The utility of encoding symmetries into machine learning algorithms has long been known (see Bronstein et al. (2021) for a modern review). With this has come a desire to learn symmetries from data, rather than encoding them directly. The classic works of Rao and Ruderman (1998) and Miao and Rao (2007) established the possibility of explicitly learning a Lie group from data, in supervised and unsupervised settings, respectively. Though there has been much subsequent interest in learning symmetries (see the discussion of works below), there has been remarkably little comparison of these methods, likely due to the lack of a quantitative metric with which to evaluate them. Instead, most papers rely on the same heuristic, manual inspection of the learned transformations. Even on the more quantitative side, such as in the appendix of the recent work of Yang et al. (2023), one only finds direct comparisons of *single generators* of Lie groups. This approach does not generalize beyond one-dimensional groups — for example, one might have two sets of generators which look completely different pairwise, but in fact generate the same group.

We seek to fill the gap above, in order to allow for rigorous evaluation of symmetry learning. To introduce the idea, suppose G and H are Lie groups acting on a dataspace X, and we want to quantify how differently they act; for example, H is the group learned by some algorithm meant to learn G. At any point $x \in X$, we can linearly approximate the actions of the groups via the actions of their Lie algebras, as shown in Figure 1. The resultant tangent spaces can be compared directly in terms of their principal angle, which is invariant under a change of basis.

In Section 3 we formally define the *local symmetry error* (LSE) at each point $x \in X$, and study some of its properties, including groups which it cannot distinguish. Though

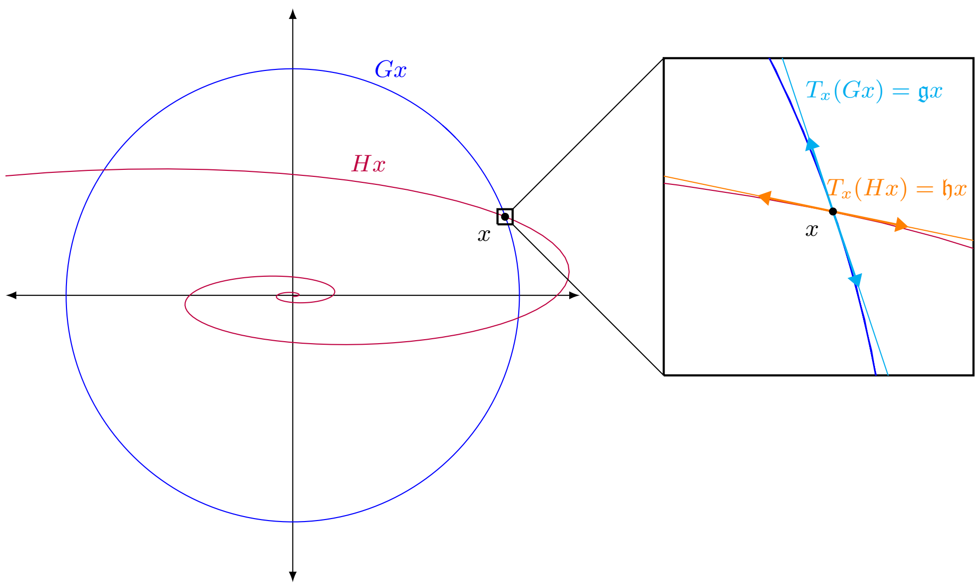

Figure 1: Under general assumptions, the orbits of a datapoint $x \in X$, under two Lie groups G and H, can be locally approximated by the tangent spaces $T_x(Gx)$ and $T_x(Hx)$. The angle between these tangent spaces can be used to quantify, locally, the difference between the actions of G and H. When the groups act by matrix multiplication on a vector space, the tangent spaces can be written as the set-wise products of the Lie algebras with the datapoint.

we focus here on the case $X = \mathbb{R}^d$ and where G, H are connected matrix Lie groups in $\mathbb{R}^{d\times d}$, the definition generalizes readily to smooth group actions on manifolds. As the LSE depends on the reference point $x \in X$, in order to obtain a single number characterizing the difference between G and H, one might average the LSE or else try to upper bound it over all $x \in X$. In Section 4 we derive a uniform upper bound on LSE, which is tight when G and H are one-dimensional groups and G is unitary. In Section 5, we demonstrate the use of LSE-based metrics by comparing quantitatively the target and learned groups for existing symmetry learning methods. The experiments we perform suggest that LSE-based metrics may be useful in identifying invariant subspaces in the data. While the current work is focused on one-dimensional groups — as a domain in which exact results are available and easily computable, and so experiments are both informative and inexpensive — it lays the groundwork for the first quantitative evaluations of symmetry learning beyond one-dimensional families. Nonetheless, the LSE as proposed here retains some problems for practical use, which we discuss in Section 6.

2. Related work

Modern works on learning symmetry can broadly be categorized into (1) works in the same vein as Rao and Ruderman (1998) and Miao and Rao (2007), seeking to learn the group itself (Cohen and Welling, 2014; Dehmamy et al., 2021; Chau et al., 2022; Desai et al., 2022; Keurti et al., 2022; Gabel et al., 2023; Yang et al., 2023), (2) works attempting to learn invariant features or filters (Wetzel et al., 2020; Wieser et al., 2020; Mouli and Ribeiro, 2021; Zhou et al., 2021; Sanborn et al., 2023), and (3) works attempting to learn the extent of invariance under a pre-specified symmetry (Benton et al., 2020; van der Ouderaa and van der Wilk, 2021; Allingham et al., 2022). The main goal of this paper is to introduce a quantitative measure of success for the first type of work, though it should in principle be possible to apply our methods for example to Sanborn et al. (2023), if one extracts the Lie algebra implied by the invariant features in the model.

Our work bears a spiritual resemblance to others which attempt to quantify the degree of symmetry learned by a supervised neural network. Moskalev et al. (2022) elaborate a method to extract infinitesimal invariances in a learned function. Gruver et al. (2023) use the Lie derivative to measure whether a learned function is locally equivariant to a desired symmetry. An important difference between these works an ours is that we are more interested in the learning of the group itself rather than an equivariant function. As such we do not require a supervised learning setting, and are able to evaluate unsupervised methods in our experiments.

3. Local Symmetry Error

Tangent spaces We will assume that $X = \mathbb{R}^d$ is the dataspace, and G, H are connected matrix Lie groups acting naturally on X. We denote by $\mathfrak{g}, \mathfrak{h}$ the groups' respective Lie algebras. For any point $x \in X$, the orbits Gx and Hx form manifolds, and we denote by $T_x(Gx)$ and $T_x(Hx)$ their respective tangent spaces at x. (See Figure 1.) Given that G, H act by matrix multiplication, one can readily show that $T_x(Gx) = \mathfrak{g}x$ and $T_x(Hx) = \mathfrak{h}x$.

Distance between subspaces In order to measure the distance between two subspaces U, V of some overarching inner product space (e.g. our tangent spaces above as subspaces of $\mathbb{R}^d$), we use the well-studied notion of their *aperture*, introduced by Krein et al. (1948),

$$\Theta(U, V) = \max \left\{ \sup_{\substack{u \in U \\ ||u||=1}} \inf_{v \in V} ||u - v||, \sup_{\substack{v \in V \\ ||v||=1}} \inf_{u \in U} ||u - v|| \right\}. \tag{1}$$

From the work cited above and Akhiezer and Glazman (1993, §34), we have the following.

Proposition 1 *Let U, V be subspaces of a Hilbert space. Then $\Theta(U, V) \in [0, 1]$, with $\Theta(U, V) < 1$ only if $\dim U = \dim V$. Assuming the dimensions are in fact equal,*

$$\Theta(U, V) = \sup_{\substack{u \in U \\ ||u||=1}} \inf_{v \in V} ||u - v|| = \sup_{\substack{v \in V \\ ||v||=1}} \inf_{u \in U} ||u - v|| \tag{2}$$

$$= ||P_U - P_V|| \tag{3}$$

$$= \sin(\varphi) \tag{4}$$

where P_U, P_V are orthogonal projectors and the norm in (3) is the operator (i.e. spectral) norm, and φ is the principal angle between U and V.

From (3) it is clear that the aperture defines an honest-to-goodness metric on subspaces of equal dimension. On the other hand, (4) gives an intuitive understanding of what the aperture is measuring.

Definition We define the *local symmetry error* (LSE) at any point $x \in X$ by

$$\mathrm{LSE}_x(G, H) = \Theta(T_x(Gx), T_x(Hx)). \tag{5}$$

In the case of matrix Lie groups, we have

$$\mathrm{LSE}_x(G, H) = \Theta(\mathfrak{g}x, \mathfrak{h}x) = ||P_{\mathfrak{g}x} - P_{\mathfrak{h}x}||, \tag{6}$$

which can be easily computed given bases for $\mathfrak{g}$ and for $\mathfrak{h}$, that is, generators of G and H.

It is worth making a few remarks. First, note that we assume here the dimensions of the tangent spaces $\mathfrak{g}x$ and $\mathfrak{h}x$ are equal. We return to the practical consequences of this point in our final discussion (Section 6). Additionally, the LSE cannot detect global differences between groups whose infinitesimal actions are the same — for example $SO(4)$ and its double-cover $SU(2) \times SU(2)$, or $SO(2) \times \mathbb{R}$ and $\mathbb{R}^2$ acting on the plane, by rotation-and-scaling and translation respectively.[1]

Since LSE is a "local" metric on Lie groups defined at any point $x \in X$, to obtain a metric which does not depend on $x \in X$ one must somehow pool the local information. In our experiments we consider two natural approaches, the *average LSE* over the data distribution $\mathbb{P}$ (though any test distribution may be considered),

$$\mathrm{ALSE}(G, H) = \mathbb{E}_{x\sim\mathbb{P}}\left[\mathrm{LSE}_x(G, H)\right], \tag{7}$$

and the *maximum LSE*,

$$\mathrm{MLSE}(G, H) = \sup_{x\in X} \mathrm{LSE}_x(G, H). \tag{8}$$

The latter is a stricter notion of similarity depending only on the spaces question, and not on the distribution from which data is actually observed.

4. Bounding Maximum Local Symmetry Error

At face value, obtaining an estimate of the MLSE (8) is non-trivial. A trivial, but nonetheless important observation is that the LSE is scale-invariant in x,

$$\mathrm{LSE}_{cx}(G, H) = ||P_{\mathfrak{g}cx} - P_{\mathfrak{h}cx}|| = ||P_{\mathfrak{g}x} - P_{\mathfrak{h}x}|| = \mathrm{LSE}_x(G, H)\ \forall c \in \mathbb{R}. \tag{9}$$

It follows that

$$\mathrm{MLSE}(G, H) = \sup_{\substack{x\in X\\ ||x||=1}} ||P_{\mathfrak{g}x} - P_{\mathfrak{h}x}||,$$

which one could in the worst case estimate by sampling a grid of unit vectors. One can alternatively upper bound the MLSE, as we do below, in terms of any subset of the Lie algebra $S \subseteq \mathfrak{g}$ for which, for any unit length $x \in X$, we have

$$\{A \in \mathfrak{g} : ||Ax|| = 1\} \subseteq S. \tag{10}$$

1. In this case, we allow ourselves to consider non-linear actions (as translation is not linear). However, it is easy to come up with similar examples using only linear actions.

If G, H are one-dimensional groups, this bound below is not only tight but easily computable if the set S is "nice."[2] We provide proofs in Appedix B.

Theorem 2 *Let G, H be connected matrix Lie groups acting naturally on $X = \mathbb{R}^d$, and suppose there exists $S \subseteq \mathfrak{g}$ such that (10) holds. Then*

$$\text{MLSE}(G, H) \leq \sup_{A \in S} \inf_{B \in \mathfrak{h}} ||A - B||. \tag{11}$$

Corollary 3 *If $G \cong U(1)$ and H are one-dimensional groups, then the inequality is an equality, from which we obtain*

$$\text{MLSE}(G, H) = \inf_{t \in \mathbb{R}} ||A_1 - tB_1|| \tag{12}$$

where $A_1 \in \mathfrak{g}$ is a matrix of unit operator norm and $B_1 \in \mathfrak{h}$ is any nonzero matrix.

Note that (12) can be computed by numerically solving a convex optimization problem. On the other hand, the upper bound in (11) is in general not easy to compute. When S is contained in the set of matrices of unit operator norm, passing to the Frobenius norm, we get a looser bound which one can appealingly express in terms of the aperture between the Lie algebras (considered as vector subspaces of $\mathbb{R}^{d \times d}$ given Frobenius inner product),

$$\frac{1}{\sqrt{d}} \Theta(\mathfrak{g}, \mathfrak{h}) \leq \text{MLSE}(G, H) \leq \sqrt{d}\, \Theta(\mathfrak{g}, \mathfrak{h}). \tag{13}$$

When both d and $\Theta(\mathfrak{g}, \mathfrak{h})$ are large, the bound above is of little use. One must therefore make a tradeoff between the looseness of the bound and the computational cost of estimating the MLSE directly in the manner mentioned previously.

5. Experiments

We demonstrate the use of LSE by calculating or estimating the ALSE (7), MLSE (8), and the aperture $\Theta(\mathfrak{g}, \mathfrak{h})$ on the symmetries learned under different settings in the work of Yang et al. (2023). Broadly, we pursue two directions.

First, we compute both exact MLSEs and empirical ALSEs for some learned one-dimensional groups (using Theorem 2 for the exact results). We observe empirically, and justify, that a discrepancy between the ALSE and MLSE indicates that the data distribution is restricted to an "interesting" subspace of the entire dataspace X. We take advantage of this setting to demonstrate how one might calculate LSE-based metrics on subspaces X, albeit in this case on a simple, linear subspace.

Second, we use the availability of exact MLSEs to provide some empirical data on the convergence of sample-based estimates of the MLSE. An understanding of this convergence is essential for good approximations of the MLSE when the group of interest is more than one-dimensional. We leave such experiments for future iterations of this work.

2. One reason we focus our experiments on this case is exactly because it allows us to explicitly compute the MLSE.

Datasets We use the code provided by Yang et al. (2023) to reproduce their experiments, which learn group generators in an unsupervised fashion on the training datasets for various tasks. We consider two settings: 2-body trajectory prediction and 3-body trajectory prediction, both in 2 dimensions. The observations consist of vectors of such as $x = (q_1, p_1, q_2, p_2)$ where q_i and p_i are 2-dimensional positions and momenta, respectively, where i indexes the body in question. In the two-body case, Yang et al. (2023) identify two valid symmetries, generated by the block matrices

$$\begin{pmatrix} R & & & \\ & R & & \\ & & R & \\ & & & R \end{pmatrix} \qquad \begin{pmatrix} R & & -R & \\ & R & & -R \\ -R & & R & \\ & -R & & R \end{pmatrix}$$

where R is the matrix for a rotation by $\pi/2$ radians. The latter symmetry holds because observations centered, so the center of mass is always at the origin. The authors thus propose three hyperparameter settings for their algorithm, which learns a group generator in $\mathbb{R}^{8\times 8}$: one in which the former block structure is enforced in the generator (without weight-tying), one in which the latter block structure is enforced (again without weight-tying), and finally one in which no structure is enforced. We reproduce all three experiments with ten separate random seeds each, and report our metrics with means and standard deviations in Table 1, denoting the hyperparameter settings by "diag," "block," and "full," respectively. In the 3-body case, the previous authors only consider the diagonal and full settings, so we follow suit. Note that in the block and full settings, we must take care which group is considered the ground truth, and whether the entire dataspace is maximized over in MLSE; we elaborate on the choices reported in the table in the next paragraph.

Table 1: LSE-based metrics for groups learned as in Yang et al. (2023)

Dataset	ALSE	MLSE (exact)	MLSE (approx.)	$\Theta(\mathfrak{g}, \mathfrak{h})$
2-body (diag)	0.009 ± 0.001	0.019 ± 0.002	—	0.012 ± 0.001
2-body (block)	0.008 ± 0.002	0.015 ± 0.003	—	0.011 ± 0.003
2-body (full)	0.010 ± 0.001	0.77 ± 0.07	—	0.47 ± 0.05
3-body (diag)	0.31 ± 0.07	0.6 ± 0.1	—	0.33 ± 0.06
3-body (full)	0.19 ± 0.04	1.000 ± 0.000	—	0.86 ± 0.06

ALSE and MLSE estimation We estimate the ALSE in all three experimental settings using the same empirical distribution from which the symmetries are learned, i.e. the training and testing datasets. For the trajectory prediction tasks, we learn one-dimensional groups and compare against representations of $SO(2)$, and can thus calculate the exact MLSEs using Theorem 2. In the "diag" setting this is straightforward. In the "block" setting, we compare against the second symmetry mentioned above.[3] However, we observe

3. Comparing against the same ground truth as "diag" of course gives an MLSE close to 1; the ALSE, however, is low, by the same reasoning as below.

that naïvely maximizing LSE over all of $\mathbb{R}^d$ in "block" setting leads to the curious results in Table 2 — namely, while the empirical ALSE is low, the MLSE is fairly high (and, correspondingly, so is the aperture). The explanation for this must be that the ALSE is restricted to a subset of datapoints at which the LSE is low. The explanation is exactly that the second symmetry identified by Yang et al. (2023) is only a symmetry on the set X_0 of datapoints $x = (q_1, p_1, q_2, p_2)$ restricted such that $q_2 = -q_1$ and $p_2 = -p_1$. This suggests that the correct MLSE should be restricted to X_0. As X_0 is linear, it is natural to try replacing all occurrences of x in Theorem 2 and its proof with $P_{X_0}x$. It is straightforward to verify this works noting it is equivalent to replacing $\mathfrak{g}$ with $\mathfrak{g}P_{X_0}$ and $\mathfrak{h}$ with $\mathfrak{h}P_{X_0}$. As one might suspect, the "full" setting behaves similarly; note that we use the "block" version of the ground truth in this setting as well. The values reported in Table 1 are with the projection inserted.

For the 3-body task, we observe in the "full" setting the same pattern as in Table 2 — relatively low ALSE and high MLSE — suggesting that the learned symmetries apply in some subspace to which the data distribution is confined. We do not attempt to identify this subspace here, but show the generator with lowest ALSE in Appendix A, Figure 2.

Table 2: LSE-based metrics on 2-body, where MLSE maximizes over all of $\mathbb{R}^d$

Dataset	ALSE	MLSE (exact)	MLSE (approx.)	$\Theta(\mathfrak{g}, \mathfrak{h})$
2-body (block)	0.008 ± 0.002	0.78 ± 0.05	—	0.71 ± 0.08

Convergence of sampling-based MLSE estimates Given the availability of exact MLSE values, we can examine how many points $x \in X$ one must sample to obtain good estimates using sample maxima. We consider sampling points from isotropic Gaussians of appropriate dimensions. We aggregate the difference between MLSE and ALSE across runs and hyperparameter settings, and plot them against the number of sampled points $x \in X$, separately for the 2-body and 3-body tasks. See Figures 3 and 4 in Appendix A. The errors appear to decrease linearly in a log-log plot. The dependence on the dimension d is of interest, but cannot be determined solely from these plots. As a rule of thumb, it appears that up to $d \approx 10$ one can get good estimates using around 10^5 samples.

6. Discussion

We have proposed a general method — as far as we can tell, the first — for evaluating symmetry learning when the objects of interest are connected Lie groups with potentially more than one generator. While this proposal is intuitively appealing, and we are able both to demonstrate its computation as a proof of concept and obtain theoretical and empirical insights, to prove truly useful it must be better understood and tested empirically. Most obviously, we shall evaluate it on multi-dimensional groups, but it is also desirable to understand other properties, such as the relationship between LSE-based metrics and the performance of equivariant models using the learned symmetries on downstream tasks. Indeed, we have begun such experiments, building on the work of Yang et al. (2023) and Finzi et al. (2021).

Additionally, the LSE has some properties which can prove inconvenient. Firstly, one might want to compare Lie algebras whose tangent spaces have different dimensions. This may be particularly important for evaluating symmetry-learning methods when the number of generators of the true group is not known. When the learned group has a different dimension than the true group, the LSE will generally be constantly 1 at all points, regardless of how "close" the learned group is to the true group. Even in the case where the dimensions of the Lie algebras match, a perturbation of the true generators may increase the dimension the tangent space at any given point $x \in X$.[4] We suspect an appropriate generalization of LSE can address both of these problems. For example, Ye and Lim (2016) present a systematic generalization of aperture to spaces of differing dimension. A second problem we identify is that the tangent space structure does not distinguish certain group actions (see our remark in Section 3). We are looking at using other information local to neighborhoods of points $x \in X$ to make such distinctions.

There is also room for further theoretical understanding, such as a better characterization (or even just examples) of when the conditions of Theorem 2 hold. Some hand-tested examples also suggest that Corollary 3 generalizes to multiple-dimensional groups; if true, this expands the practicality of the MLSE as an explicitly computable metric.

Acknowledgments

The author thanks Eric Sager Luxenberg for helpful discussions on minimax theorems.

References

N.I. Akhiezer and I.M. Glazman. *Theory of Linear Operators in Hilbert Space*. Dover Books on Mathematics. Dover Publications, 1993. ISBN 9780486677484. URL https://books.google.co.uk/books?id=9-xfAK8xOqMC.

James Urquhart Allingham, Javier Antoran, Shreyas Padhy, Eric Nalisnick, and José Miguel Hernández-Lobato. Learning generative models with invariance to symmetries. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022. URL https://openreview.net/forum?id=Ff1N3et1IV.

Edgar Asplund and Vlastimil Pták. A minimax inequality for operators and a related numerical range. *Acta Mathematica*, 126(none):53 – 62, 1971. doi: 10.1007/BF02392025. URL https://doi.org/10.1007/BF02392025.

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew G Wilson. Learning invariances in neural networks from training data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17605–17616. Curran Associates, Inc., 2020.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021.

4. Consider for example perturbing the generators of $SO(3)$.

Ho Yin Chau, Frank Yuchen Qiu, Yubei Chen, and Bruno Olshausen. Disentangling images with lie group transformations and sparse coding. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022. URL https://openreview.net/forum?id=SZEL_mvz7eU.

Taco Cohen and Max Welling. Learning the irreducible representations of commutative lie groups. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1755–1763, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/cohen14.html.

Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 2503–2515. Curran Associates, Inc., 2021.

Krish Desai, Benjamin Nachman, and Jesse Thaler. Symmetry discovery with deep learning. *Phys. Rev. D*, 105:096031, 5 2022.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3318–3328. PMLR, 7 2021.

Alex Gabel, Victoria Klein, Riccardo Valperga, Jeroen S. W. Lamb, Kevin Webster, Rick Quax, and Efstratios Gavves. Learning lie group symmetry transformations with neural networks, 2023.

Nate Gruver, Marc Anton Finzi, Micah Goldblum, and Andrew Gordon Wilson. The lie derivative for measuring learned equivariance. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=JL7Va5Vy15J.

Hamza Keurti, Hsiao-Ru Pan, Michel Besserve, Benjamin F Grewe, and Bernhard Schölkopf. Homomorphism autoencoder — learning group structured representations from interactions. In *UAI 2022 Workshop on Causal Representation Learning*, 2022. URL https://openreview.net/forum?id=9XUM3-KJ50U.

M.G. Krein, M. A. Krasnoseljsky, and D. P. Milman. On the defect indices of linear operators in banach space and on some geometric questions. In *Sbornik Trudov Instituta Matematiki Akad. Nauk UkrSSR*, 1948. URL https://apps.dtic.mil/sti/pdfs/AD0663073.pdf.

Xu Miao and Rajesh P. N. Rao. Learning the Lie Groups of Visual Invariance. *Neural Computation*, 19(10):2665–2693, 10 2007. ISSN 0899-7667. doi: 10.1162/neco.2007.19.10.2665. URL https://doi.org/10.1162/neco.2007.19.10.2665.

Artem Moskalev, Anna Sepliarskaia, Ivan Sosnovik, and Arnold W.M. Smeulders. LieGG: Studying learned lie group generators. 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=9sKZ60VtRmi.

S Chandra Mouli and Bruno Ribeiro. Neural networks for learning counterfactual g-invariances from single environments. In *International Conference on Learning Representations*, 2021.

Rajesh Rao and Daniel Ruderman. Learning lie groups for invariant visual perception. In M. Kearns, S. Solla, and D. Cohn, editors, *Advances in Neural Information Processing Systems*, volume 11. MIT Press, 1998. URL https://proceedings.neurips.cc/paper_files/paper/1998/file/277281aada22045c03945dcb2ca6f2ec-Paper.pdf.

Sophia Sanborn, Christian A Shewmake, Bruno Olshausen, and Christopher J. Hillar. Bispectral neural networks. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=xnsg4pfKb7.

Tycho van der Ouderaa and Mark van der Wilk. Learning invariant weights in neural networks. In *ICML 2021 Workshop on Uncertainty & Robustness in Deep Learning.* 2021.

Sebastian J. Wetzel, Roger G. Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Phys. Rev. Research*, 2:033499, 9 2020.

Mario Wieser, Sonali Parbhoo, Aleksander Wieczorek, and Volker Roth. Inverse learning of symmetries. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18004–18015. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/d15426b9c324676610fbb01360473ed8-Paper.pdf.

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 39488–39508. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/yang23n.html.

Ke Ye and Lek-Heng Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197, 2016. doi: 10.1137/15M1054201. URL https://doi.org/10.1137/15M1054201.

Allan Zhou, Tom Knowles, and Chelsea Finn. Meta-learning symmetries by reparameterization. In *International Conference on Learning Representations*, 2021.

Appendix A. Additional figures

The figures referenced in Section 5 are included below.

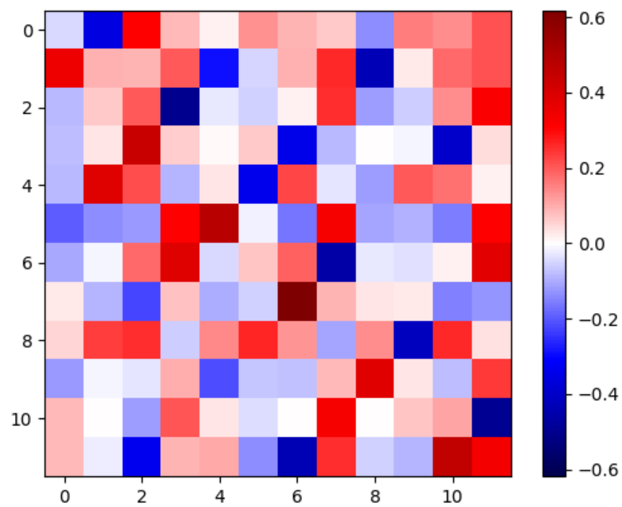

Figure 2: Generator with lowest ALSE in 3-body "full" setting

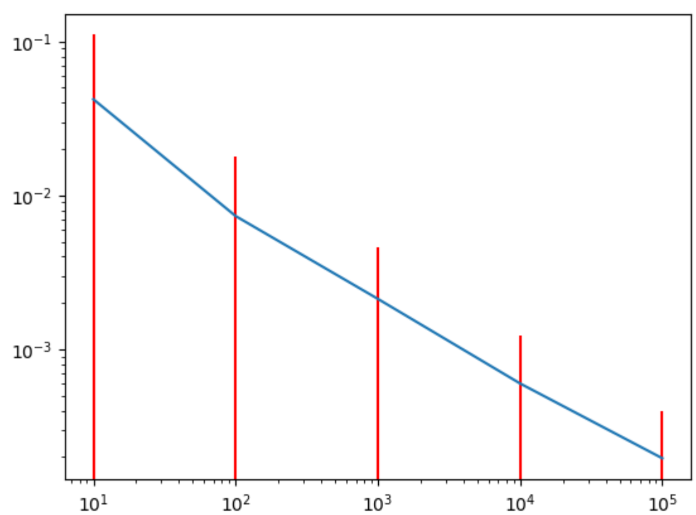

Figure 3: Error in MLSE estimate against number of random Gaussian datapoints, 2-body task (log-log axes, standard deviation in red)

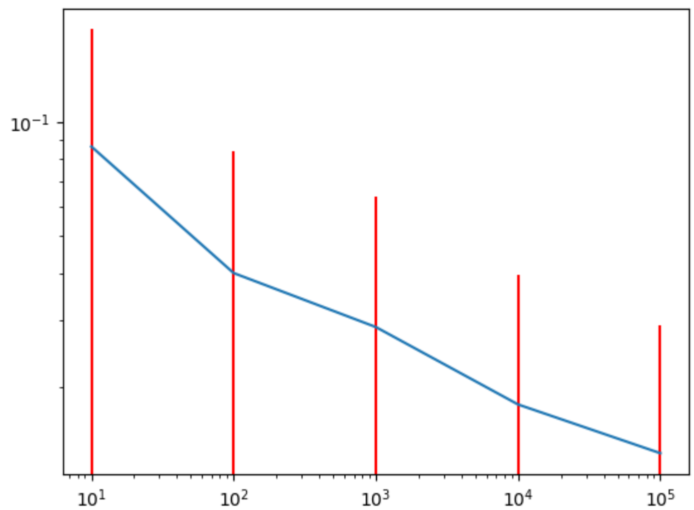

Figure 4: Error in MLSE estimate against number of random Gaussian datapoints, 2-body task (log-log axes, standard deviation in red)

Appendix B. Proofs

B.1. Proof of Theorem 2

Using Proposition 1 (2), we can write

$$\mathrm{MLSE}(G,H) = \sup_{\substack{x\in X\\ ||x||=1}} \sup_{\substack{A\in\mathfrak{g}\\ ||Ax||=1}} \inf_{B\in\mathfrak{h}} ||Ax - Bx||.$$

By assumption, the above is upper bounded as

$$\mathrm{MLSE}(G,H) \leq \sup_{A\in S} \sup_{\substack{x\in X\\ ||x||=1}} \inf_{B\in\mathfrak{h}} ||Ax - Bx||. \tag{14}$$

A standard result on exchanging suprema and infima finishes the proof

$$\mathrm{MLSE}(G,H) \leq \sup_{A\in S} \inf_{B\in\mathfrak{h}} \sup_{\substack{x\in X\\ ||x||=1}} ||Ax - Bx||.$$

B.2. Proof of Corollary 3

When $G \cong SO(2)$, the Lie algebra $\mathfrak{g}$ is the span of a single skew-symmetric matrix. The matrix R corresponding to a $\pi/2$-radian rotation is one of two elements of $\mathfrak{g}$ with unit operator norm (the other being its negative). One can then verify that for any $x \in X$ with unit norm,

$$\{A \in \mathfrak{g} : ||Ax|| = 1\} = \{R, -R\} = S,$$

for example by noting that by the skew symmetry of elements $A \in \mathfrak{g}$, the product $Ax \cdot Ax = -x \cdot A^2 x$ can only be 1 if x is an eigenvector of A^2 with eigenvalue -1. Thus the inequality in (14) is in fact an equality. The final interchange of the inner supremum and infimum maintains equality in this case because $\mathfrak{h}$ is one-dimensional, applying the minimax theorem of Asplund and Pták (1971).