Self-Supervised Latent Symmetry Discovery via Class-Pose Decomposition

Gustaf Tegnér GUSTAFTE@KTH.SE
Hedvig Kjellström HEDVIG@KTH.SE
KTH Royal Institute of Technology, Stockholm

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

Abstract

In this paper, we explore the discovery of latent symmetries of data in a self-supervised manner. By considering sequences of observations undergoing uniform motion, we can extract a shared group transformation from the latent observations. In contrast to previous work, we utilize a latent space in which the group and orbit component are decomposed. We show that this construction facilitates more accurate identification of the properties of the underlying group, which consequently results in an improved performance on a set of sequential prediction tasks.

Keywords: Representation Theory, Representation Learning, Group Theory, Symmetry Discovery

1. Introduction

Symmetry discovery poses a fundamental challenge in geometric deep learning and has been explored in a range of previous work (Rao and Ruderman (1998); Cohen and Welling (2014); Yang et al. (2023); Dehmamy et al. (2021)). In this paper, we consider the problem of discovering *latent* symmetries from high-dimensional data. This entails both finding a representation, and the group acting upon it in a simultaneous manner. Learning latent representations naturally induces a problem of identifiability, as there can be a possibly infinite set of representations that are equally valid. We ask the question to what extent one can identify the single ground-truth underlying group structure from high-dimensional observations.

In Meta-Sequential Prediction (MSP) (Miyato et al. (2022)), latent symmetries are learned by considering a self-supervised forward prediction task on sequences arising from uniform motion. An example would be predicting the next frame from a video of objects undergoing 3D rotation and translation. The assumption of uniform motion leads to a problem setup where you have a shared group action across distinct pairs of data points. From this, a latent transformation can be learnt in a self-supervised manner that can accurately depict the latent transitions. MSP learns accurate transitions, but suffer on long-term prediction, as their latent representation entangles the pose of an object and its class, resulting in redundancy in the group representation. This has consequences on both the predictive performance, but also on the identification of the correct group. In this work, we expand upon this setting by inducing the latent representation with a structure that distinguishes the pose of the object from its class. Such a representation has previously been considered by Marchetti et al. (2023), although we extend it to the case where the group structure

is unknown. We show that under certain conditions, this leads to a representation that is conjugate to the true group, entailing that spectral properties such as its eigenvalues are preserved. Furthermore, we showcase our proposed method on a sequential prediction task for a number of groups, achieving more accurate long-term predictions compared to MSP.

2. Background

We consider a space $\mathcal{X}$ which on which a group G *acts* via the map $(g, x) = g \circ x$. The set of elements that are connected by a group action induces an equivalence class $\mathcal{X} \setminus G$ which we denote as the *orbits* of G on $\mathcal{X}$. As an example, the space $\mathcal{X}$ could consist of images of a set of objects under different orientations. An objects pose can be expressed by its orientation from a fixed reference point. Similarly, a rotated object remains the same object, which is to say that the orbits remain invariant to group actions. Given a high-dimensional dataset such as images, we wish to discover the group of symmetries acting upon it, or which transformations the objects in the images are undergoing. These transformations can be *non-linear* in the image plane, such as 3D rotations of the underlying object itself. This distinguishes latent symmetries from the symmetries incorporated into G-equivariant neural networks, which consider symmetries on the pixel grid (Cohen and Welling (2016)). As shown in Marchetti et al. (2023), there exists an equivariant isomorphism $\mathcal{X} \cong G \times \mathcal{X} \setminus G$. This implies that a possibly high-dimensional space $\mathcal{X}$ can be completely encoded with a low-dimensional group structure, without loss of information. Learning the group then entails finding a representation of $\mathcal{X}$ and G that induces this isomorphism.

2.1. Meta-Sequential Prediction

Meta-Sequential Prediction (Miyato et al. (2022)) presents a method to extract symmetries by considering time-series of objects undergoing motion of constant velocity. To this end, they consider trajectories $\mathbf{x}_g = \{x_t = g^{t-1} \circ x_1 \in \mathcal{X}, t \in [1, n]\}$ constructed through repeated application of the same group action to an initial state x_1. Different actions induce different sequences and they consider the collection of this, $\{\mathbf{x}_g, g \in G\}$, as their dataset. They introduce a representation learner $\varphi : \mathcal{X} \to \mathcal{Z}$ parameterized as a neural network with $\mathcal{Z} = \mathbb{R}^{m \times d}$. The group is assumed as a subgroup $G \leq GL(m)$ of the group of $m \times m$ invertible matrices, acting upon the latent space through matrix multiplication. For each sequence, MSP aims to learn the latent group action M such that it is *equivariant* w.r.t. the representation φ:

$$M\varphi(x) = \varphi(g \circ x). \tag{1}$$

Thus, acting in the data space should correspond to an equivalent action in the latent space.

2.2. Learning Equivariance

To learn the equivariant map, MSP utilizes an equivariance loss defined over a trajectory $\mathbf{x}$ as:

$$\mathcal{L}_{\text{equiv}}(\varphi, M, \mathbf{x}) = \frac{1}{n-1} \sum_{t=2}^{n} d(M\varphi(x_{t-1}), \varphi(x_t)) \tag{2}$$

where d is some appropriate metric, in this case the L_2 metric. The key observation in learning the equivariance is that a subsequence $\mathbf{x_S} = (x_1, \ldots, x_{n_s})$ shares the same transformation as the subsequent observations $\mathbf{x_Q} = (x_{n_s}, \ldots, x_{n_s+n_q})$ with $n = n_s + n_q$. Denote $\mathbf{x}_S$ and $\mathbf{x}_Q$ as the *support* and *query*-set respectively, borrowing notions from meta-learning literature (Finn et al. (2017)). This setup lends itself to a bi-level optimization scheme where the group transformation is learnt from the support-set, while the representations are learnt on the query-set. To this end, the optimization objective can be expressed as

$$\underset{\varphi}{\text{argmin}} \quad \mathbb{E}_{\mathbf{x_S}, \mathbf{x_Q} \sim \mathcal{X}} \left[\mathcal{L}_{\text{equiv}}(\varphi, M^*, \mathbf{x}_Q) \right] \tag{3}$$

$$\text{subject to} \quad M^* = \underset{M}{\text{argmin}}\, \mathcal{L}_{\text{equiv}}(\varphi, M, \mathbf{x}_S) \tag{4}$$

Since M^* is the solution to a linear system of equations, it can be found in closed form. In addition, to ensure injectivity, a further reconstruction loss is implemented through a decoder $\psi : \mathcal{Z} \to \mathcal{X}$.

3. Method

In this section we introduce a simple extension to MSP to ensure better identifiability of the group components. MSP embeds $\mathcal{X} = G \times \mathcal{X} \setminus G$ into a latent space $\mathcal{Z} = \mathbb{R}^{m \times d}$. As $M \in \mathbb{R}^{m \times m}$ acts *linearly* on this representation, the orbit must remain an invariant. To enable this, MSP encodes the orbit into a subspace that is invariant to transformations of M. This is only possible if M is *reducible*, i.e. there exists a G-invariant subspace that is non-trivial. To enable this, the dimension of M must be strictly greater than the dimension of the group. Encoding into higher dimensions, however, entails identifying more spurious representations.

To circumvent this entanglement, we propose a representation that decomposes class (orbit) from pose (group). To this end, we propose to set the latent space as $\mathcal{Z} = G \times \mathcal{E}$ which represents a decomposition of the latent space into the group and orbit. We represent G as a subgroup of $GL(m)$ and the orbit component as $\mathcal{E} \subseteq \mathbb{R}^d$. We decompose our encoder $\varphi = (\varphi_G, \varphi_\mathcal{E})$ as the group and orbit encoder respectively. To train φ_G, we consider the objective as defined in Equation 3. Similarly $\varphi_\mathcal{E}$ is trained to be invariant within a sequence by enforcing all its encodings to be equal. Similarly to MSP, we also train a decoder ψ for reconstruction. As M is implicitly dependent on the group action g, we can consider it a map from $M : G \to GL(m)$. As proven in Miyato et al. (2022), M is a group homomorphism and thus a representation of G. As φ is injective, it follows that M is a faithful representation of G. Thus, when restricted to its orbits, it induces an isomorphism between G and its image $M(G)$. In certain cases, this image is isomorphic to a unique group representation. For simple and semi-simple groups, any faithful representation of minimal dimension is irreducible (Joyce, 2000, chapter 3). In example, embedding $SO(3)$ in $GL(3)$ represents an embedding of minimal dimension and thus, the representation will be irreducible. Furthermore, for $SO(3)$, there exists a unique irreducible representation (Hall and Hall, 2013, chapter 4). Thus we can make a stronger claim, that the representation learnt will be *conjugate* to $SO(3)$, and preserve its spectral properties.

4. Experiments

We demonstrate our proposed method on a set of synthetic regressions tasks. We construct sequences (x_t) by sampling an initial condition and group action uniformly from G and letting $x_t = g^{t-1} \circ x_1$ for $t \in [1, T]$. We construct orbits by sampling a projection vector $P_o \in \mathbb{R}^{m^2 \times D}$ and projecting each x into $\mathbb{R}^D$. We consider the rotational groups $SO(n)$ for $n \in \{2, 3\}$. For each group, we generate a dataset of 1000 different initial conditions and group actions and generate sequences of length $T = n_s + n_q$ with $n_s = 10$ and $n_q = 1$.

4.1. Experimental Details

We implement our model as a 3-layer neural network with 64 hidden units and ReLU activations. We utilize Batch Normalization (Ioffe and Szegedy (2015)) applied after each activation function as this significantly reduced the training time. We train the models for 200 epochs using the Adam Optimizer (Kingma and Ba (2014)) and a learning rate of 0.001. The results are presented as the average over 5 random seeds evaluated on a test dataset representing 10% of the data. Our orbit encoder $\varphi_{\mathcal{E}}$ encodes to a 16 dimensional latent space, which is the same dimension we use for dimension d in MSP.

4.2. Results

In Table 1 we present results of the average rollout error over 20 time-steps. This exemplifies the generalization performance of the learnt transition beyond the 1-step training objective. When considering a single orbit, the models performances match, as they differ only in the number of extra dimensions in the encoding. However, as we increase the number of orbits, the rollout prediction error drastically increases for MSP. We can relate this to the fact that the representational capacity of MSP under a limited amount of dimensions is inadequate to encode both group and orbit. In Figure 4.2, we present measurements of the determinant of M^* during training with 5 orbits. As is shown, our method preserves the determinant while MSP fails to capture the properties of the group. Additionally, utilizing a higher-dimensional latent space fails to capture this inherent property of the group.

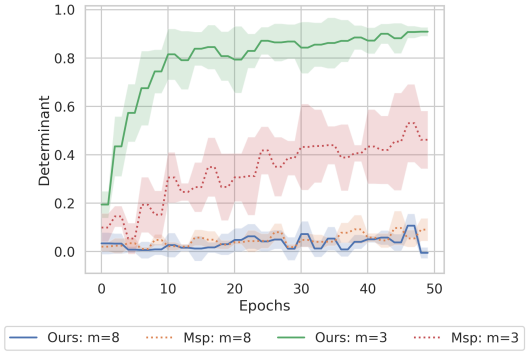

Figure 1: Determinant of M^* in $SO(3)$ considering 5 orbits.

5. Future Work

In this work, we have learnt non-linear symmetries by considering data undergoing uniform motion. Another form of symmetry emerges when the motion is possibly non-linear, but is ultimately equivariant to a group G. An example would be the dynamics of a particle system which remains equivariant to the Euclidean group (Satorras et al. (2021)). As the

	SO(2)			SO(3)		
Model / Orbits	1	10	20	1	10	20
MSP, $m = \dim(G)$	$0.96_{\pm 0.04}$	$11.63_{\pm 5.21}$	$60.73_{\pm 17.19}$	$6.74_{\pm 0.75}$	$61.88_{\pm 7.46}$	$83.00_{\pm 2.40}$
Ours, $m = \dim(G)$	$0.71_{\pm 0.06}$	$\mathbf{1.87_{\pm 0.36}}$	$4.16_{\pm 0.71}$	$5.44_{\pm 0.75}$	$\mathbf{7.59_{\pm 0.39}}$	$24.80_{\pm 4.28}$
MSP, $m = 8$	$\mathbf{0.64_{\pm 0.09}}$	$3.50_{\pm 0.40}$	$7.27_{\pm 0.14}$	$8.80_{\pm 0.61}$	$22.02_{\pm 0.43}$	$31.05_{\pm 2.86}$
Ours, $m = 8$	$0.67_{\pm 0.06}$	$2.37_{\pm 0.29}$	$\mathbf{3.18_{\pm 0.22}}$	$\mathbf{5.23_{\pm 1.24}}$	$11.54_{\pm 0.29}$	$\mathbf{20.30_{\pm 2.22}}$

Table 1: MSE ($\times 10^{-2}$) for rollouts of length 20 for a different number of orbits. As we increase the number of orbits, MSP gradually loses its ability to model the correct transformations.

dynamics and symmetry have to be learnt simultaneously, it naturally lends itself to the framework we have studied and provides an avenue for future work. Furthermore, we have shown that identifiability is possible on the simple groups we have considered. Future work could consider more complex groups, which however, may require further restrictions on the group representation.

Acknowledgements

We would like to acknowledge Giovanni Luca Marchetti for insightful discussions. This work was supported by the Swedish Research Council, Knut and Alice Wallenberg Foundation and the European Research Council (ERC-BIRD-884807).

References

Taco Cohen and Max Welling. Learning the irreducible representations of commutative lie groups. In *International Conference on Machine Learning*, pages 1755–1763. PMLR, 2014.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. *Advances in Neural Information Processing Systems*, 34:2503–2515, 2021.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

Brian C Hall and Brian C Hall. *Lie groups, Lie algebras, and representations*. Springer, 2013.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

William Peter Joyce. *Formulation of the Racah-Wigner calculus using category theory.* University of Canterbury. Physics, 2000.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Giovanni Luca Marchetti, Gustaf Tegnér, Anastasiia Varava, and Danica Kragic. Equivariant representation learning via class-pose decomposition. In *International Conference on Artificial Intelligence and Statistics*, pages 4745–4756. PMLR, 2023.

Takeru Miyato, Masanori Koyama, and Kenji Fukumizu. Unsupervised learning of equivariant structure from sequences. *Advances in Neural Information Processing Systems*, 35: 768–781, 2022.

Rajesh Rao and Daniel Ruderman. Learning lie groups for invariant visual perception. *Advances in neural information processing systems*, 11, 1998.

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

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. *arXiv preprint arXiv:2302.00236*, 2023.

	SO(2)			SO(3)		
Model / Orbits	1	3	5	1	3	5
MSP $m = \dim(G)$	$1.01_{\pm 0.06}$	$21.12_{\pm 1.94}$	$19.73_{\pm 12.05}$	$1.66_{\pm 0.05}$	$19.40_{\pm 11.76}$	$18.10_{\pm 7.18}$
Ours $= \dim(G)$	$0.95_{\pm 0.06}$	$\mathbf{7.58_{\pm 4.35}}$	$\mathbf{1.73_{\pm 0.36}}$	$1.84_{\pm 0.06}$	$\mathbf{11.41_{\pm 12.88}}$	$\mathbf{4.14_{\pm 1.67}}$
MSP $m = 8$	$\mathbf{0.65_{\pm 0.04}}$	$24.03_{\pm 8.01}$	$20.86_{\pm 5.18}$	$\mathbf{1.56_{\pm 0.06}}$	$26.74_{\pm 7.59}$	$34.25_{\pm 3.39}$
Ours $m = 8$	$0.71_{\pm 0.05}$	$23.83_{\pm 3.80}$	$18.28_{\pm 2.68}$	$1.82_{\pm 0.03}$	$23.65_{\pm 8.75}$	$39.79_{\pm 6.04}$

Table 2: MSE ($\times 10^{-2}$) of next-state prediction using the learnt transformation from a different orbit.

Appendix A. Appendix

A.1. Additional Experiments

We consider a simplified version of the dataset where orbits are generated by adding an integer $z_o \in \{0, \dots, N_{\text{orbits}}\}$ to the sequences. In this setting, we find that we can achieve *full equivariance* implying that the learnt M transfers across different orbits. We measure this by finding M^* for one orbit, and using it to evaluate the equivariance loss in Equation 2 on a sequence from another orbit. We present the results in Table 2. For one orbit, the results are mostly equivalent, with MSP showing a slight advantage. However, as the number of orbits increase, our learnt transitions showcase stronger generalization across orbits. The results imply that a complete disentanglement between class and pose is possible in certain conditions, although there is nothing explicit in the method that encourages this. Disentanglement, in this case, remains a consequence of the implicit regularization of neural networks, where a structured latent space may encourage it, but not explicitly impose it.