SE(3) FRAME EQUIVARIANCE IN DYNAMICS MODELING AND REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

In this paper, we aim to explore the potential of symmetries in improving the understanding of continuous control tasks in the 3D environment, such as locomotion. The existing work in reinforcement learning on symmetry focuses on pixel-level symmetries in 2D environments or is limited to value-based planning. Instead, we considers continuous state and action spaces and continuous symmetry groups, focusing on translational and rotational symmetries. We propose a pipeline to use these symmetries in learning dynamics and control, with the goal of exploiting the underlying symmetry structure to improve dynamics modeling and model-based planning.

1 MOTIVATION

Symmetries have recently been explicitly included into some machine learning algorithms, such as translational equivariance of convolution neural networks (CNNs) on image segmentation. Symmetries can provide regularities from the laws of physics, including physics-related tasks such as modeling dynamics. However, it has not been widely explored in the context of control of dynamical system, such as continuous control tasks in reinforcement learning.

In exiting work, van der Pol et al. (2020b); Mondal et al. (2020) initiate the study of symmetry in deep reinforcement learning, after (Ravindran and Barto; Zinkevich and Balch, 2001). They focus on pixel-level symmetries in the CartPole task or Atari games, such as left-right reflection or discrete rotations. Nevertheless, such level of symmetry does not unveil the symmetry of the underlying dynamical system. Zhao et al. (2022b) study discrete symmetry in model-based planning on 2D discrete grid, which uses the symmetry of 2D grid (4 rotations and 2 reflections). However, it is constrained to 2D grid and value-based planning, which cannot trivially extend to continuous control case that needs sampling-based planning.

This motivates to further move ahead to continuous state and action spaces and continuous symmetry groups and ask: *can we make use of the symmetry structure of the 3D space to improve dynamics modeling and model-based planning/control?* In this work, we aim to exploit the underlying symmetry structure of those continuous control tasks, through the lens of symmetry. We focus on the tasks in the 3D environment, which include locomotion and manipulation and simulate the needs of our 3D physical world. We consider translational and rotational symmetries in the 3D space, which are related to conservation of linear and angular momentum. To this end, we propose a pipeline to use these symmetries in learning dynamics and control.

2 RELATED WORK

Symmetries widely exist in various domains and have been exploited in classic planning algorithms and model checking (Fox and Long, 1999; 2002; Pochter et al., 2011; Domshlak et al.; Shleyfman et al., 2015; Sievers et al., 2015; Sievers; Sievers et al., 2019; Fier et al., 2019). Zinkevich and Balch (2001) show the invariance of value function for an MDP with symmetry. However, these algorithms have a fundamental issue with the exploitation of symmetries, as they explicitly construct equivalence classes of symmetric states, which are intractable (NP-hard) in maintaining symmetries in trajectory rollout and forward search (Narayanamurthy and Ravindran, 2008) and incompatible with differentiable pipelines for representation learning. To address this issue, recent work has explored

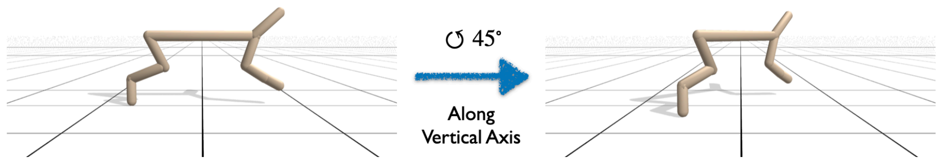

Figure 1: A half cheetah is learning to walk on 2D surface in a 3D physical environment. Rotating the half cheetah (along vertical z-axis), the dynamics is equivariant, and the optimal behaviors do not change.

state abstraction for symmetry, such as coarsest state abstraction that aggregates symmetric states into equivalence classes, studied in MDP homomorphisms and bisimulation (Ravindran and Barto, 2004; Ferns et al., 2004; Li et al., 2006). However, these methods usually require perfect MDP dynamics knowledge and do not scale well due to the complexity of constructing and maintaining abstraction mappings (van der Pol et al., 2020a). Several recent studies have integrated symmetry into RL based on MDP homomorphisms (Ravindran and Barto, 2004). van der Pol et al. (2020a) integrate symmetry by equivariant policy network, which avoids the difficulties in handling symmetry in forward search. Earlier, Mondal et al. (2020) have separately applied similar idea earlier while not using MDP homomorphisms. Park et al. (2022) learn equivariant transition models, but do not consider planning. Zhao et al. (2022a) focus on permutation symmetry in object-oriented transition models. The direct precedent is (Zhao et al., 2022b), which studies 2D discrete symmetry on 2D grid with a value-based planning approach.

3 Problem Formulation

3.1 Graph Representation of a Dynamical System

We consider a dynamical system that models a robot in a 3D physical world. One example is locomotion, where a robot learns to move on a ground that provides support, with gravity directed downwards. The robot is represented by connected links (bodies) and joints (actuators). Suppose the system is modeled by a discrete-time Markov decision process (MDP) as $s' = f(s, a)$. As done in Interaction Network (Battaglia et al., 2016) or Graph Network (Sanchez-Gonzalez et al., 2018), we could represent the robot state s at each time step as a geometric graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

This geometric graph lives in a 3D geometric space, allowing for rotation and translation transformations. The graph nodes $v_i \in \mathcal{V}$ are links with features of positions, orientations, linear velocity, and angular velocity. The edges $e_{ij} \in \mathcal{E}$ describe the joints between links, where actuation is provided as edge feature. As an example, we can rotate the half cheetah $45°$ along the vertical axis, where all nodes and edges are rotated correspondingly, as shown in the figure.

The frame symmetry in 3D is the focus of our interest. It implies that even if we change the reference frame by moving in different directions or locations, the robot does not need to relearn how to walk or run. The frame symmetry represents the proper transformations of the entire 3D world, represented as $\mathrm{SE}(3)$, the group of 3D continuous translations and rotations. However, the robot is subject to external forces that may break symmetry, such as gravity, contact, and actuation forces. Despite these forces, the $\mathrm{SE}(3)$ symmetry can still be maintained if they are included as inputs

3.2 Symmetry in Continuous Control

In this section, we outlines the symmetry under consideration in the continuous control tasks and how we explore that. (Ravindran and Barto, 2004; Ravindran and Barto; Zinkevich and Balch, 2001) explore symmetry in MDPs with no function approximation like neural networks. (van der Pol et al., 2020a; Mondal et al., 2020) initiate the exploration of symmetry in model-free (deep) RL by using equivariant policy networks. (Zhao et al., 2022b) study symmetry in value-based planning on 2D grid. We extend it and focus on MDPs with continuous state and action spaces and sampling-based control/planning algorithms.

Symmetry properties. The *symmetry* properties in MDPs are specified by equivariance of the transition and reward functions, studied in Zinkevich and Balch (2001); Ravindran and Barto (2004);

van der Pol et al. (2020a); Zhao et al. (2022b):

$$\bar{P}(s' \mid s, a) = \bar{P}(g \cdot s' \mid g \cdot s, g \cdot a), \quad \forall g \in G, \forall s, a, s' \tag{1}$$

$$\bar{R}_M(s, a) = \bar{R}_{g \cdot M}(g \cdot s, g \cdot a), \quad \forall g \in G, \forall s, a \tag{2}$$

Note that how the group G *acts* on states and actions is called *group representation*, and is decided by the space $\mathcal{S}$ or $\mathcal{A}$. Zhao et al. (2022b) study path planning and take maps M as a part of input. On high level, this takes into consideration all "symmetry breaking" factors, such as obstacles in path planning that are represented by input occupancy maps M. In typical continuous control tasks, such as locomotion and manipulation, the symmetry breaking factors include external forces, such as gravity, contact (e.g., with ground), and actuation forces (from control motors).

4 Frame Equivariance in Dynamics Model and Control

Our high-level objective is to investigate the advantages of symmetry structures in dynamic systems for enhancing continuous control and planning. Additionally, we seek to minimize the complexity of task-specific design for symmetry, reducing the barriers to implementing equivariant methods and avoiding any unnecessary burdens.

Overview. When symmetry presents in the dynamical system, the optimal policy and value functions are also equivariant (Ravindran and Barto, 2004; Ravindran and Barto; Zinkevich and Balch, 2001; van der Pol et al., 2020a). The existence of symmetry helps shrink the hypothesis spaces and reduced generalization gap. In this section, we propose an approach to make use of symmetry in sampling-based planning methods, such as model predictive control (MPC).

Our formulation is an extension to the prior work on symmetric planning (Zhao et al., 2022b), which designs a principled approach to consider discrete symmetry. Specifically, they focus on path planning in 2D discrete grids ($\mathbb{Z}^2$) with a discrete symmetry group (D_4) consisting of rotations and reflections for value-based planning. The key insight is that all functions defined on $\mathbb{Z}^2$ signals ($x : \mathbb{Z}^2 \to \mathbb{R}^d$) are *steerable* by the D_4 group. Since every step is equivariant, the entire value iteration process $\texttt{VI}(M)$ is equivariant:

$$g.\texttt{VI}(M) \equiv g.\mathcal{T}^\infty[V_0] = \mathcal{T}^\infty[g.V_0] \equiv \texttt{VI}(g.M), \tag{3}$$

where M is the occupancy map and goal input $\mathbb{Z}^2 \to \{0, 1\}^2$. Additionally, equivariant mappings between these $\mathbb{Z}^2$ signals can be represented as convolutions and other operations (e.g., $+$, $\times$, $\max$), which can be implemented using CNNs. Our work generalizes to 3D continuous space $\mathbb{R}^3$ and the continuous symmetry group $\mathrm{SE}(3)$. The robot acts in continuous action space, which necessitates the consideration of sampling-based method.

4.1 Dynamics Model with Equivariance

To use frame equivariance in control, there are three key steps (Zhao et al., 2022b): (1) specify the symmetry group G in the system, (2) learn a G-equivariant dynamics model $P(s' \mid s, a)$, and (3) incorporate symmetry into continuous control. In the second step, using equivariant model induces equivariant Bellman operator: $\mathcal{T}[V] = \max_a R(s, a) + \gamma \sum_{s'} P(s'|s, a)V(s')$. Albeit the previous two steps are analogous to prior work, the third step is not trivial because of the nature of sampling-based approaches. We explain it in the next subsection.

Since we use graph-based representation of the system, we make use of equivariant message passing networks for implementing the dynamics network. Specifically, we use steerable $\mathrm{E}(3)$-equivariant message passing network (Brandstetter et al., 2022). Compared to using invariant *scalar* features, steerable networks allow to use higher-order equivariant *steerable* features, enabling better expressivity. We briefly explain how we use steerable message passing network and specify $\mathrm{SO}(3)$-representations for state and action features. We omit the details on hidden layers and more.

Group acting on geometric graphs. In order to understand how the group acts on the states and actions represented by geometric graphs $\mathcal{G}$, we discuss the group representations of $\mathrm{SE}(3)$ rotations and translations. We focus on the representations of $\mathrm{SO}(3)$ rotations, $\rho(g), g \in \mathrm{SO}(3)$. We say a vector $\boldsymbol{h}$ is *steerable* if there exists a matrix $\mathbf{D}$ that can transform $\boldsymbol{h}$ for group elements g by $\mathbf{D}(g)\boldsymbol{h}$. In short, the representations of $\mathrm{SO}(3)$ can be decomposed into irreducible representations, or Wigner-D matrices $\mathbf{D}^{(l)}(g)$ with dimensions $(2l+1) \times (2l+1)$. We say a vector $\boldsymbol{h}$ transformed by l-th matrix as *type-l steerable vector*. A type-0 steerable feature is a scalar (trivial representation), and a type-1 feature is steerable by 3×3 rotation matrices (standard representation).

The graph nodes are links with features of positions (3D, type-1 feature), orientations (quaternion, $\mathrm{SO}(3)$ acts by group composition), linear velocity (3D, type-1 feature), and angular velocity (also 3D vector space, tangent space, type-1 feature). The edge features include actions, which are typically scalars or type-0 features. We also include gravity direction $\boldsymbol{g}$ as a type-1 global feature, which is also transformed correspondingly. We also include additional features, computed from quantities like positions and velocities, similar to (Brandstetter et al., 2022).

4.2 Control with Symmetry

In prior work (Zhao et al., 2022b), they developed Symmetric Value Iteration Network (SymVIN) for path planning on 2D grid $\mathbb{Z}^2$. Intuitively, SymVIN relate maps under four discrete rotations and two reflections, so if a map is rotated, it does not need to relearn the optimal plan or actions. Analogously, we aim to develop a control method that considers the frame equivariance of 3D space. One practical setup is locomotion on some terrain. The symmetry of locomotion task is that a robot does not to learn how to walk when facing different directions and locations, while instead share information between all directions and locations of $\mathrm{SE}(3)$.

Symmetry in MPC. Our goal is to exploit SE(3) frame symmetry of the underlying $\mathbb{R}^3$ space in the planning algorithm. For sampling-based planning, symmetries enable two types of benefits.

1. *Forward search.* When sampling a trajectory $(s_1, a_1, s_2, a_2, \ldots)$ in the ground MDP, we equivariantly know the outcome of all trajectories under symmetries: $\{(g \cdot s_1, g \cdot a_1, g \cdot s_2, g \cdot a_2, \ldots) \mid g \in G\}$. This saves computation when the group is "large". However, while this is important for discrete case, it is negligible for continuous state space.
2. *Value backup.* Symmetries allow us to reuse knowledge between equivalent state, as $V^\star(s) = V^\star(g \cdot s)$ and $\pi(g \cdot s) = g \cdot \pi(s)$. Intuitively, facing different directions do not change how a robot walks.

In continuous control, the second type of consideration is crucial. Since the optimal policy and value functions are equivariant (or invariant), we constrain the function approximation to be only the set of equivariant functions.

Planning with MPC. We use model predictive control as a sampling-based planning approach for continuous actions. We use MPPI (Model Predictive Path Integral) control method (Williams et al., 2015; 2016; 2017a;b), as also done in (Hansen et al., 2022). We sample N trajectories with horizon H using the learned dynamics model with actions from a learned policy, and estimate the expectation of total return G_τ. Importantly, if we rotate the entire trajectory τ by $g \cdot \tau$, we know the return is invariant $G_\tau = G_{g \cdot \tau}$ since it is a scalar:

$$G_\tau \triangleq \mathbb{E}_\tau\left[\gamma^H Q_\theta\left(\boldsymbol{s}_H, \boldsymbol{a}_H\right) + \sum_{t=0}^{H-1} \gamma^t R_\theta\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right)\right] \tag{4}$$

$$= \mathbb{E}_{g \cdot \tau}\left[\gamma^H Q_\theta\left(g \cdot \boldsymbol{s}_H, g \cdot \boldsymbol{a}_H\right) + \sum_{t=0}^{H-1} \gamma^t R_\theta\left(g \cdot \boldsymbol{s}_t, g \cdot \boldsymbol{a}_t\right)\right] \tag{5}$$

We can then update the action selection policy by using top-k trajectories, which is a Gaussian distribution with learned mean and variance. To capture the symmetry here, we use equivariant policy network and invariant value network in estimating values and optimal actions.

5 Discussion

In this work, we provide a principled guideline to consider the continuous $\mathrm{SE}(3)$ frame symmetry in sampling-based planning and control in 3D environments, which extends the prior work of path planning with value-based planning on 2D grid with 2D discrete symmetry. As a working project, we are still working on tuning results for equivariant dynamics and symmetric control methods.

REFERENCES

P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, and K. Kavukcuoglu. Interaction Networks for Learning about Objects, Relations and Physics. *arXiv:1612.00222 [cs]*, Dec. 2016. URL http://arxiv.org/abs/1612.00222. arXiv: 1612.00222.

J. Brandstetter, R. Hesselink, E. van der Pol, E. J. Bekkers, and M. Welling. Geometric and Physical Quantities Improve E(3) Equivariant Message Passing. *arXiv:2110.02905 [cs, stat]*, Mar. 2022. URL http://arxiv.org/abs/2110.02905. arXiv: 2110.02905.

C. Domshlak, M. Katz, and A. Shleyfman. Enhanced Symmetry Breaking in Cost-Optimal Planning as Forward Search. page 5. doi: 10/gq5m5r.

N. Ferns, P. Panangaden, and D. Precup. Metrics for Finite Markov Decision Processes. In *AAAI*, 2004.

D. Fier, . Torralba, and A. Shleyfman. Operator Mutexes and Symmetries for Simplifying Planning Tasks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):7586–7593, July 2019. ISSN 2374-3468. doi: 10/ghkkbq. URL https://ojs.aaai.org/index.php/AAAI/article/view/4751. Number: 01.

M. Fox and D. Long. The Detection and Exploitation of Symmetry in Planning Problems. In *In IJCAI*, pages 956–961. Morgan Kaufmann, 1999.

M. Fox and D. Long. Extending the exploitation of symmetries in planning. In *In Proceedings of AIPS02*, pages 83–91, 2002.

N. Hansen, X. Wang, and H. Su. Temporal Difference Learning for Model Predictive Control. Technical Report arXiv:2203.04955, arXiv, Mar. 2022. URL http://arxiv.org/abs/2203.04955. arXiv:2203.04955 [cs] type: article.

L. Li, T. J. Walsh, and M. Littman. Towards a Unified Theory of State Abstraction for MDPs. In *AI&M*, 2006.

A. K. Mondal, P. Nair, and K. Siddiqi. Group Equivariant Deep Reinforcement Learning. *arXiv:2007.03437 [cs, stat]*, June 2020. URL http://arxiv.org/abs/2007.03437. arXiv: 2007.03437.

S. M. Narayanamurthy and B. Ravindran. On the hardness of finding symmetries in Markov decision processes. In *Proceedings of the 25th international conference on Machine learning - ICML '08*, pages 688–695, Helsinki, Finland, 2008. ACM Press. ISBN 978-1-60558-205-4. doi: 10/bkswc2. URL http://portal.acm.org/citation.cfm?doid=1390156.1390243.

J. Y. Park, O. Biza, L. Zhao, J. W. van de Meent, and R. Walters. Learning Symmetric Embeddings for Equivariant World Models. *arXiv:2204.11371 [cs]*, Apr. 2022. URL http://arxiv.org/abs/2204.11371. arXiv: 2204.11371.

N. Pochter, A. Zohar, and J. S. Rosenschein. Exploiting Problem Symmetries in State-Based Planners. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, Aug. 2011. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3732.

B. Ravindran and A. G. Barto. Symmetries and Model Minimization in Markov Deision Proesses.

B. Ravindran and A. G. Barto. *An algebraic approach to abstraction in reinforcement learning*. PhD thesis, University of Massachusetts at Amherst, 2004.

A. Sanchez-Gonzalez, N. Heess, J. T. Springenberg, J. Merel, M. Riedmiller, R. Hadsell, and P. Battaglia. Graph networks as learnable physics engines for inference and control. *arXiv:1806.01242 [cs, stat]*, June 2018. URL http://arxiv.org/abs/1806.01242. arXiv: 1806.01242.

A. Shleyfman, M. Katz, M. Helmert, S. Sievers, and M. Wehrle. Heuristics and Symmetries in Classical Planning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1), Mar. 2015. ISSN 2374-3468. doi: 10/gq5m5s. URL https://ojs.aaai.org/index.php/AAAI/article/view/9649. Number: 1.

S. Sievers. Structural Symmetries of the Lifted Representation of Classical Planning Tasks. page 8.

S. Sievers, M. Wehrle, M. Helmert, and M. Katz. An Empirical Case Study on Symmetry Handling in Cost-Optimal Planning as Heuristic Search. In S. Hlldobler, R. Pealoza, and S. Rudolph, editors, *KI 2015: Advances in Artificial Intelligence*, volume 9324, pages 166–180. Springer International Publishing, Cham, 2015. ISBN 978-3-319-24488-4 978-3-319-24489-1. doi: 10.1007/978-3-319-24489-1_13. URL http://link.springer.com/10.1007/978-3-319-24489-1_13. Series Title: Lecture Notes in Computer Science.

S. Sievers, G. Rger, M. Wehrle, and M. Katz. Theoretical Foundations for Structural Symmetries of Lifted PDDL Tasks. *Proceedings of the International Conference on Automated Planning and Scheduling*, 29:446–454, 2019. ISSN 2334-0843. doi: 10/gq5m5t. URL https://ojs.aaai.org/index.php/ICAPS/article/view/3509.

E. van der Pol, D. Worrall, H. van Hoof, F. Oliehoek, and M. Welling. Mdp homomorphic networks: Group symmetries in reinforcement learning. *Advances in Neural Information Processing Systems*, 33, 2020a.

E. van der Pol, D. E. Worrall, H. van Hoof, F. A. Oliehoek, and M. Welling. MDP Homomorphic Networks: Group Symmetries in Reinforcement Learning. *arXiv:2006.16908 [cs, stat]*, June 2020b. URL http://arxiv.org/abs/2006.16908. arXiv: 2006.16908.

G. Williams, A. Aldrich, and E. Theodorou. Model Predictive Path Integral Control using Covariance Variable Importance Sampling. *arXiv:1509.01149 [cs]*, Oct. 2015. URL http://arxiv.org/abs/1509.01149. arXiv: 1509.01149.

G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou. Aggressive driving with model predictive path integral control. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1433–1440, May 2016. doi: 10/gf9knc.

G. Williams, A. Aldrich, and E. A. Theodorou. Model Predictive Path Integral Control: From Theory to Parallel Computation. *Journal of Guidance, Control, and Dynamics*, 40(2):344–357, Feb. 2017a. ISSN 0731-5090, 1533-3884. doi: 10/f9vx74. URL https://arc.aiaa.org/doi/10.2514/1.G001921.

G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, and E. A. Theodorou. Information theoretic MPC for model-based reinforcement learning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1714–1721, Singapore, May 2017b. IEEE. ISBN 978-1-5090-4633-1. doi: 10/ggdv8n. URL https://ieeexplore.ieee.org/document/7989202/.

L. Zhao, L. Kong, R. Walters, and L. L. S. Wong. Toward Compositional Generalization in Object-Oriented World Modeling. In *ICML 2022*, Apr. 2022a. URL http://arxiv.org/abs/2204.13661. arXiv: 2204.13661.

L. Zhao, X. Zhu, L. Kong, R. Walters, and L. L. S. Wong. Integrating Symmetry into Differentiable Planning. In *ICLR 2023*. ICLR, June 2022b. doi: 10.48550/arXiv.2206.03674. URL http://arxiv.org/abs/2206.03674. arXiv:2206.03674 [cs] type: article.

M. Zinkevich and T. Balch. Symmetry in Markov decision processes and its implications for single agent and multi agent learning. In *In Proceedings of the 18th International Conference on Machine Learning*, pages 632–640. Morgan Kaufmann, 2001.