

Lie Group Error Coordinates for Symmetry-Aware Reinforcement Learning applied to Quadrotor Low-Level Control

Andrea Pagnini and Ezio Malis

Abstract—As data-driven methods become prevalent in robotics, a key question remains whether classical geometric structures are still relevant or whether they can be learned from data. We argue that geometry is not an alternative to learning, but a design tool that shapes what must be learned. In this paper, we show that encoding the right symmetry in the observation of a RL agent reduces the effective complexity of the control problem at the representation level, prior to any architectural choice. We demonstrate this principle on quadrotor low-level control, expressing tracking errors as Lie group quantities in the desired body frame. We show that this coordinate choice improves sample efficiency and enables zero-shot generalization to unseen trajectories, suggesting that the right choice of error coordinates can effectively improve learning without relying on architectural changes.

I. INTRODUCTION

Geometric deep learning [1], [2] has emerged as a powerful strategy to improve sample efficiency in learning algorithms. When considering Reinforcement Learning (RL) algorithms [3], a central question at the intersection of geometric mechanics and data-driven control is whether a system’s symmetries must be encoded in the architecture of a learning system [4], or in its representation. Two prior works of Yu and Lee on quadrotor control exploit the vehicle’s $SO(3)_{\mathbf{e}_3}$ rotational symmetry about the gravity axis \mathbf{e}_3 to improve data efficiency of the RL algorithm. In [5], symmetry is imposed by canonicalization, rotating the absolute state to a fixed representation of its orbit before feeding a standard Multilayer Perceptron (MLP), reducing input dimension by one. In [6], the learning bias is moved into the architecture, encoding the same $SO(3)_{\mathbf{e}_3}$ symmetry in Equivariant MLP (EMLP) layers [7]. Following [5], Welde et al. [8] extend symmetry reduction to stochastic trajectory tracking Markov Decision Processes (MDPs), exploiting the natural Lie group symmetries of free-flying robotic systems to construct MDP homomorphisms, enabling a policy trained on a smaller quotient MDP to be lifted to an optimal controller for the original system. Moving to the field of geometric control, Hampsey et al. [9] demonstrate that expressing the tracking error as Lie group quantities in the desired frame yields more structured error dynamics and improved LQR tracking without changing controller complexity. This demonstrates that the right choice of symmetry and error representation is fundamental to shape the structure of the resulting error dynamics, and consequently, the control problem.

This work has been supported by the ASCAR project, funded by French National Research Agency (ANR-23-ASTR-0016).

The authors are with ACENTAURI team at Centre Inria d’Universite Côte d’Azur, Sophia-Antipolis, France, {andrea.pagnini, ezio.malis}@inria.fr

We transfer this principle to RL, expressing Lie group tracking errors in the desired body frame. Our contributions are: (i) an $SO(3)$ -invariant representation that makes group actions trivial for standard MLPs, removing architectural constraints; and (ii) simulative evidence of improved sample efficiency and zero-shot generalization to unseen trajectories.

II. EQUIVARIANT RL FOR QUADROTOR CONTROL

A. Quadrotor UAV Modeling

Let \mathcal{W} and \mathcal{B} be the world and body frames, with \mathbf{z}_W pointing up, \mathbf{x}_B pointing forward and \mathbf{z}_B aligned with the total thrust. The state vector $\mathbf{x} = (\mathbf{p}, \mathbf{v}, \mathbf{R}, \boldsymbol{\omega})$ is composed of position in the world frame, linear velocity in the world frame, rotation from body to world frame, and angular velocity in the body frame, respectively. The system evolves on the manifold $\mathcal{M} = \mathbb{R}^3 \times \mathbb{R}^3 \times SO(3) \times \mathbb{R}^3$ according to:

$$\begin{aligned} \dot{\mathbf{p}} &= \mathbf{v}, & \dot{\mathbf{v}} &= \frac{F}{m} \mathbf{R} \mathbf{e}_3 - g \mathbf{e}_3, \\ \dot{\mathbf{R}} &= \mathbf{R} \hat{\boldsymbol{\omega}}, & \dot{\boldsymbol{\omega}} &= \mathbf{J}^{-1}(\boldsymbol{\tau} - \boldsymbol{\omega} \times \mathbf{J} \boldsymbol{\omega}), \end{aligned} \quad (1)$$

where m is the mass, \mathbf{J} the inertia matrix, g the gravity constant, and $\mathbf{e}_3 = [0, 0, 1]^\top$. The operator $\hat{\cdot}$ denotes the standard $\mathfrak{so}(3)$ isomorphism. Control inputs $\mathbf{u} = [F, \boldsymbol{\tau}^\top]^\top$ consist of total thrust F and body torque $\boldsymbol{\tau}$, mapped from rotor forces via a standard allocation matrix [10].

B. Equivariant Reinforcement Learning

Consider the Markov Decision Process (MDP) defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \gamma)$, where \mathcal{S} is the state space, \mathcal{A} is the action space, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ the reward function, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{P}(\mathcal{S})$ is the transition probability distribution, and $\gamma \in (0, 1)$ is the discount factor for future rewards. Suppose the system’s state evolves following a deterministic equation of motion $\dot{s} = f(s, a)$, with $f : \mathcal{S} \times \mathcal{A} \rightarrow T\mathcal{S}$. The value function associated with a policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$ is:

$$V_\pi(t, s(t)) = \int_t^\infty \gamma^{\tau-t} \mathcal{R}(s(\tau), \pi(s(\tau))) d\tau. \quad (2)$$

The goal of a RL algorithm is to identify the optimal policy π^* that maximizes the value function $V_\pi(t, s(t))$. As the sample efficiency of model-free RL is usually low, and data acquisition is expensive, Yu and Lee [6] propose to leverage symmetries in robot’s dynamics by designing an equivariant policy architecture. Let G be a Lie group acting on the state manifold via ϕ_g and on the action manifold via ψ_g [11]. A map $h : \mathcal{M} \rightarrow \mathcal{N}$ is G -equivariant if it commutes with the respective group actions, and G -invariant if the action on the codomain is the identity. Specifically, the dynamics f are equivariant if $f(\phi_g(s), \psi_g(a)) = d\phi_g(f(s, a))$, where

$d\phi_g$ is the differential of the state action. [6, Prop 4] proves that if the dynamics f are equivariant and the reward \mathcal{R} is invariant, then the optimal value function V^* is G -invariant and the optimal policy π^* is G -equivariant, i.e., $\pi^*(\phi_g(s)) = \psi_g(\pi^*(s))$. This result permits restricting the learning domain to the quotient space $(\mathcal{S} \times \mathcal{A})/\sim$, where \sim represents the equivalence class $[s, a] = \{(\phi_g(s), \psi_g(a)) \mid g \in G\}$. By enforcing these symmetry constraints via Equivariant Neural Networks [7], we reduce the effective dimensionality of the problem by $\dim(G)$, significantly improving sample efficiency of the learning process. This approach is applied to quadrotor-low level control, where the equations of motion in (1) are equivariant with respect to the group $G = SO(3)_{\mathbf{e}_3}$ of rotations about the gravity axis \mathbf{e}_3 . The state and input group actions are defined as $\phi_{\mathbf{Q}}(\mathbf{x}) = (\mathbf{Q}\mathbf{p}, \mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{R}, \boldsymbol{\omega})$ and $\psi_{\mathbf{Q}}(\mathbf{u}) = (F, \boldsymbol{\tau})$ for any $\mathbf{Q} \in SO(3)_{\mathbf{e}_3}$, respectively. Body-frame quantities are invariant under the group action. [6] defines the observation vector $\mathbf{o}_{nom} = (\boldsymbol{\varepsilon}_p, \boldsymbol{\varepsilon}_v, \mathbf{R}, \boldsymbol{\varepsilon}_{b_1}, \boldsymbol{\varepsilon}_\omega) \in \mathbb{R}^{10} \times SO(3)$, and the reward function

$$\mathcal{R}_{nom} = -k_p \|\boldsymbol{\varepsilon}_p\|^2 - k_v \|\boldsymbol{\varepsilon}_v\|^2 - k_{b_1} |\boldsymbol{\varepsilon}_{b_1}| - k_\omega \|\boldsymbol{\varepsilon}_\omega\|^2 - r_{crash} \quad (3)$$

where each $k \in \mathbb{R}^+$ is a positive constant, $\boldsymbol{\varepsilon}_p = \mathbf{p} - \mathbf{p}_d$, $\boldsymbol{\varepsilon}_v = \mathbf{v} - \mathbf{v}_d$ and $\boldsymbol{\varepsilon}_\omega = \boldsymbol{\omega} - \boldsymbol{\omega}_d$ are respectively the position, velocity and angular velocity errors, while r_{crash} is a large positive constant added to the reward when the quadrotor crashes. $\boldsymbol{\varepsilon}_{b_1}$ is the scalar heading error, defined as the angle between the desired yaw projected onto the plane normal to the thrust direction. Considering the group action acting simultaneously on the current and desired states \mathbf{x} and \mathbf{x}_d , the observation and its dynamics are equivariant, as $\mathbf{o}_{nom}(\phi_{\mathbf{Q}}(\mathbf{x}), \phi_{\mathbf{Q}}(\mathbf{x}_d)) = \phi_{\mathbf{Q}}(\mathbf{o}_{nom}(\mathbf{x}, \mathbf{x}_d))$ and $\dot{\mathbf{o}}_{nom}(\phi_{\mathbf{Q}}(\mathbf{x}), \phi_{\mathbf{Q}}(\mathbf{x}_d)) = d\phi_{\mathbf{Q}}(\dot{\mathbf{o}}_{nom}(\mathbf{x}, \mathbf{x}_d))$. Under the same observation definition, the reward is invariant, as $\mathcal{R}_{nom}(\phi_{\mathbf{Q}}(\mathbf{x}), \phi_{\mathbf{Q}}(\mathbf{x}_d)) = \mathcal{R}_{nom}(\mathbf{x}, \mathbf{x}_d)$, thanks to the norm operation. This allows the design of an equivariant policy architecture, reducing the effective dimensionality of the problem by one.

III. LIE GROUP ERROR COORDINATES

In this section, we show that by expressing the tracking error as a Lie group quantity in the desired body frame, it is possible to obtain an invariant error dynamics, by embedding the full $SO(3)$ symmetry directly into the observation, and thus allowing a standard MLP to exploit it without architectural constraints. We define the error as:

$$\mathbf{E} = \underbrace{(\mathbf{R}_d^T(\mathbf{p} - \mathbf{p}_d))}_{\mathbf{E}_p}, \underbrace{\mathbf{R}_d^T(\mathbf{v} - \mathbf{v}_d)}_{\mathbf{E}_v}, \underbrace{\mathbf{R}_d^T \mathbf{R}}_{\mathbf{E}_R}, \underbrace{\boldsymbol{\omega} - \mathbf{R}^T \mathbf{R}_d \boldsymbol{\omega}_d}_{\mathbf{E}_\omega}, \quad (4)$$

where \mathbf{E}_p , \mathbf{E}_v , and \mathbf{E}_R follow the desired body frame error formulation in [9], while \mathbf{E}_ω is computed as in [12]. This error representation is invariant under any simultaneous rotation $\mathbf{Q} \in SO(3)$ of the current and desired states, i.e. $\mathbf{E}(\phi_{\mathbf{Q}}(\mathbf{x}), \phi_{\mathbf{Q}}(\mathbf{x}_d)) = \mathbf{E}(\mathbf{x}, \mathbf{x}_d)$, where the left group action is defined as $\phi_{\mathbf{Q}}(s) = (\mathbf{Q}\mathbf{p}, \mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{R}, \boldsymbol{\omega})$ for any $\mathbf{Q} \in SO(3)$. It follows that any state on the same $SO(3)$ orbit of the state-reference pair $(\mathbf{x}, \mathbf{x}_d)$ maps to the same error

coordinates, embedding the symmetry directly into the error representation, and thus simplifying the learning problem structure without relying on Equivariant Neural Networks. The invariance property above implies also invariance of the error dynamics, as $\dot{\mathbf{E}}(\phi_{\mathbf{Q}}(s), \phi_{\mathbf{Q}}(s_d)) = \dot{\mathbf{E}}(s, s_d)$, where:

$$\begin{aligned} \dot{\mathbf{E}}_p &= -\hat{\boldsymbol{\omega}}_d \mathbf{E}_p + \mathbf{E}_v, & \dot{\mathbf{E}}_v &= -\hat{\boldsymbol{\omega}}_d \mathbf{E}_v + \frac{F}{m} \mathbf{E}_R \mathbf{e}_3 - \frac{F_d}{m} \mathbf{e}_3, \\ \dot{\mathbf{E}}_R &= \mathbf{E}_R \hat{\mathbf{E}}_\omega, & \dot{\mathbf{E}}_\omega &= \dot{\boldsymbol{\omega}} + \hat{\mathbf{E}}_\omega \mathbf{E}_R^T \boldsymbol{\omega}_d - \mathbf{E}_R^T \dot{\boldsymbol{\omega}}_d. \end{aligned} \quad (5)$$

Proof comes from the fact that the error dynamics is function only of the error and desired body frame quantities, which are both invariant by construction. Considering near-hover conditions allows to simplify the above equations by fixing $\boldsymbol{\omega}_d = 0$, $\dot{\boldsymbol{\omega}}_d = 0$ and $F_d = mg$, leading to the observation:

$$\mathbf{o}_{err} = (\mathbf{E}_p, \mathbf{E}_v, \mathbf{E}_R, \boldsymbol{\varepsilon}_{b_1}, \mathbf{E}_\omega) \in \mathbb{R}^{10} \times SO(3). \quad (6)$$

This simplification allows to preserve the same $SO(3)$ invariance property, and obtain an observation vector with the same dimension as the one in [6], required to provide a fair comparison. It is worth noting that while the input dimension to the RL algorithm is the same in the two approaches, the effective dimensionality is 2 dimensions lower in our error formulation, as the group action is a generic $SO(3)$ rotation and not just a fixed axis rotation. The reward function is the same defined in (3), which is also invariant under the same group action, as rotation does not affect the norm.

IV. SIMULATIONS

In order to validate the proposed error-coordinate formulation, we compare four different methods using the same actor-critic architecture utilizing two-layer networks, where critic networks feature 62 hidden nodes and actor networks consist of two layers with 24 nodes. Nom-MLP corresponds to the standard setup with observation \mathbf{o}_{nom} , while Nom-EMLP corresponds to the same observation but with an EMLP architecture, encoding the $SO(3)_{\mathbf{e}_3}$ symmetry about the gravity axis as proposed in [6]. Err-MLP and Err-EMLP correspond to the proposed error-coordinate observation \mathbf{o}_{err} , when using a standard MLP and an EMLP architecture, respectively. The 4 methods are compared across three actor-critic algorithms: PPO [13], SAC [14] and TD3 [15]. We follow the training setup of [6], where initial states are sampled uniformly from a 1m^3 box around the origin, and domain randomization perturbs physical parameters by $\pm 10\%$ per episode. Fig. 1 shows the learning curves averaged over four different initial training seeds, where the mean of the Average Return is reported with a shaded region representing the standard deviation. When comparing error-coordinate policies, we can notice that using a standard MLP or an equivariant EMLP architecture leads to similar performances, confirming that the symmetry reduction is already attained at observation level and does not require further architectural modification. In this sense, Err-EMLP is redundant in implementing the symmetry reduction. Err-MLP matches or accelerates the convergence of Nom-EMLP across all algorithms, with the advantage being most pronounced under TD3 and SAC, where 90% of peak return is

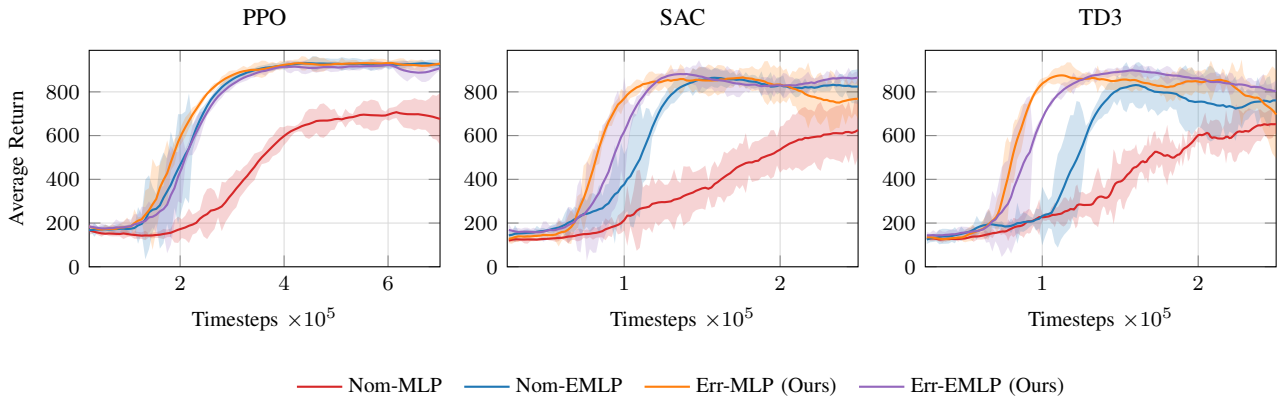


Fig. 1: Learning curves over different RL algorithms for Nom-MLP (red), Nom-EMLP (blue), Err-MLP (orange) and Err-EMLP (purple). Each curve is the mean of the Average Return over 4 seeds, shaded regions represent standard deviation.

reached in about 35% and 25% fewer timesteps, respectively. Both are off-policy algorithms that train from a replay buffer populated throughout the entire training history. Err-MLP’s invariant coordinates ensure that every buffered transition looks identical regardless of when it was collected, reducing distributional shift across the buffer. TD3 further amplifies this effect due to its deterministic nature. PPO, being on-policy, trains exclusively on freshly collected rollouts and does not accumulate such buffer inconsistencies, providing no gain for Err-MLP over Nom-EMLP. It is also worth noting that since the setpoint stabilization task considers a null pitch and roll desired angles, the $SO(3)$ symmetry of the error coordinates is not fully excited during training, this reduces to a practical $SO(3)_{e_3}$ symmetry, similar to the one in [6]. Training on aggressive trajectory tracking tasks that excite the full $SO(3)$ symmetry is left for future work and is expected to further increase the advantage of the proposed error formulation. After analyzing the learning curves, we test the capability of the trained policies to generalize over different tasks. Table I and II report the RMSE of the position \bar{e}_p and heading \bar{e}_{b_1} errors, along with the Success Rate SR , for the setpoint stabilization and Figure-8 tracking tasks, respectively. A trial is considered successful if the quadrotor does not crash, and statistics are computed over 25 evaluations for each of the 4 seeds, for a total of 100 evaluations per method. The first task is the same as considered during training, starting from a random initial state. As expected Nom-EMLP and Err-MLP show similar performance on this task, reflecting the same trend shown in the learning curves. On the other hand, Err-MLP provides a significantly higher generalization capability in the second task, when tracking a Lissajous Figure-8 trajectory [6], unseen during training, initialized from random states. In this zero-shot setting, Err-MLP achieves 100% success rate across all algorithms, with consistently lower RMSE with respect to Nom-EMLP.

V. CONCLUSIONS AND FUTURE WORK

We have shown that expressing tracking errors as Lie group quantities in the desired body frame embeds a full $SO(3)$ invariance directly into the observation, making the

TABLE I: Setpoint stabilisation performance (mean \pm std over 25 evaluations and 4 seeds for 100 total trials). Bold denotes the best value per metric for each algorithm.

Algo.	Method	\bar{e}_p [cm]↓	\bar{e}_{b_1} [deg]↓	SR [%]↑
PPO	Nom-EMLP	12.94 \pm 4.39	3.71 \pm 2.69	100
	Err-MLP (Ours)	12.61 \pm 4.34	3.11 \pm 2.03	100
SAC	Nom-EMLP	12.46 \pm 3.93	3.76 \pm 2.88	100
	Err-MLP (Ours)	11.26 \pm 3.60	2.93 \pm 1.54	100
TD3	Nom-EMLP	12.57 \pm 3.91	3.98 \pm 3.10	100
	Err-MLP (Ours)	12.42 \pm 4.11	2.94 \pm 1.54	100

TABLE II: Figure-8 tracking performance (mean \pm std over 25 evaluations and 4 seeds for 100 total trials). Bold denotes the best value per metric for each algorithm. Models are trained on setpoint stabilization.

Algo.	Method	\bar{e}_p [cm]↓	\bar{e}_{b_1} [deg]↓	SR [%]↑
PPO	Nom-EMLP	34.22 \pm 8.83	4.66 \pm 3.80	100.0
	Err-MLP (Ours)	24.76 \pm 5.09	2.71 \pm 1.72	100.0
SAC	Nom-EMLP	33.48 \pm 8.19	4.42 \pm 3.70	90.0
	Err-MLP (Ours)	25.49 \pm 4.81	2.73 \pm 1.71	100.0
TD3	Nom-EMLP	34.04 \pm 8.96	4.54 \pm 3.71	100.0
	Err-MLP (Ours)	25.03 \pm 5.15	2.83 \pm 2.04	100.0

group action trivial in the error space and allowing a standard MLP to exploit it without architectural constraints. Beyond matching or exceeding the sample efficiency of EMLP-based approaches, this coordinate choice yields a structural generalization capability, as demonstrated by zero-shot Figure-8 tracking. These results suggest a broader design principle: the right error coordinates resolve symmetry at the representation level, prior to any architectural decision, freeing equivariant architectures to potentially encode other geometric structure in the error dynamics rather than compensating for a sub-optimal observation space. The current evaluation excites only a practical $SO(3)_{e_3}$ symmetry due to the near-hover training task. Training on aggressive trajectories with non-trivial desired attitudes is the critical next step. Future work will also include real-hardware validation via sim-to-real transfer and integration with modular control architectures.

REFERENCES

- [1] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, 2017.
- [2] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *International conference on machine learning*. PMLR, 2016.
- [3] R. S. Sutton, A. G. Barto, et al., *Reinforcement learning: An introduction*. MIT press Cambridge, 1998.
- [4] 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*, 2020.
- [5] B. Yu and T. Lee, "Equivariant reinforcement learning for quadrotor uav," in *IEEE American Control Conference (ACC)*, 2023.
- [6] —, "Equivariant reinforcement learning frameworks for quadrotor low-level control," *IEEE Transactions on Control Systems Technology*, 2025.
- [7] M. Finzi, M. Welling, and A. G. Wilson, "A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups," in *International conference on machine learning*. PMLR, 2021.
- [8] J. Welde, N. Rao, P. Kunapuli, D. Jayaraman, and V. Kumar, "Leveraging symmetry to accelerate learning of trajectory tracking controllers for free-flying robotic systems," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025.
- [9] M. Hampsey, P. van Goor, T. Hamel, and R. Mahony, "Exploiting different symmetries for trajectory tracking control with application to quadrotors," *IFAC-PapersOnLine*, 2023.
- [10] J. C. Pereira, V. J. S. Leite, and G. V. Raffo, "Nonlinear model predictive control on SE(3) for quadrotor aggressive maneuvers," *Journal of Intelligent & Robotic Systems*, 2021.
- [11] J. M. Lee, *Introduction to Smooth Manifolds*. Springer, 2013.
- [12] T. Lee, M. Leok, and N. H. McClamroch, "Geometric tracking control of a quadrotor UAV on SE(3)," in *IEEE Conference on Decision and Control (CDC)*, 2010.
- [13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [14] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. Pmlr, 2018.
- [15] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*. PMLR, 2018.