Robust Reinforcement Learning Differential Game Guidance

Summary and Key Results

Ali Bani Asad

1 Research Summary

This work develops a zero-sum multi-agent reinforcement learning (MARL) framework for robust low-thrust spacecraft guidance in the Earth–Moon Circular Restricted Three-Body Problem (CR3BP). The guidance problem is posed as a differential game between a spacecraft controller (guidance agent) and an adversarial disturbance agent that injects worst-case uncertainties such as sensor noise, actuator disturbances, time delays, model mismatch, and initial-state errors.

Four state-of-the-art continuous-control algorithms (DDPG, TD3, SAC, PPO) are extended to two-player zero-sum variants and trained in a centralized-training, decentralized-execution (CTDE) setting. Extensive Monte Carlo evaluations across multiple uncertainty scenarios show that the zero-sum policies significantly improve trajectory tracking, fuel efficiency, and robustness compared with both classical control baselines and standard single-agent RL. In combined uncertainty tests over 1000 episodes, the best method (MA–TD3) reduces trajectory error by roughly 30% and improves the success rate from 88.2% (TD3) to 95.3%, while remaining suitable for real-time on-board deployment via a C++ inference stack and ROS 2 integration.

2 Performance Summary

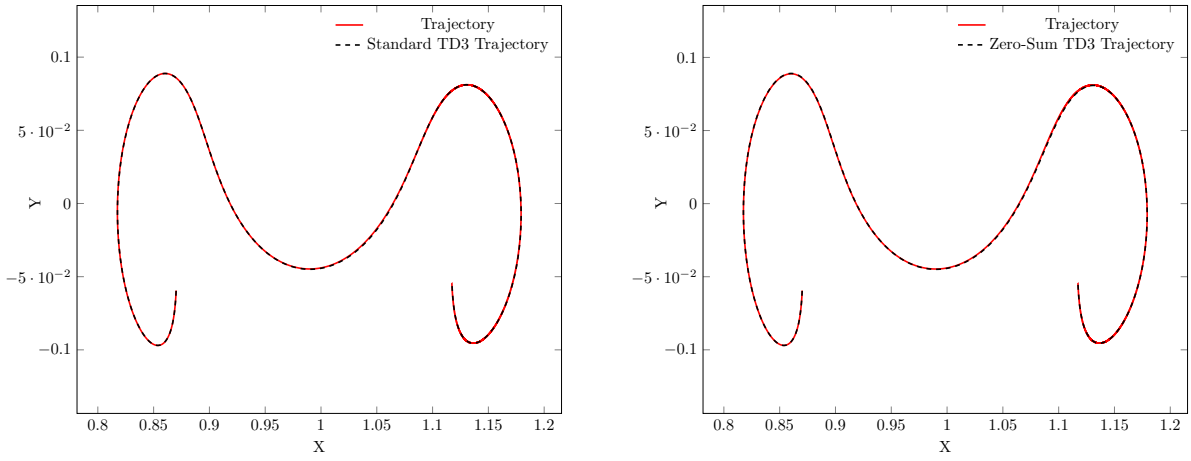
Table 1 summarizes the main quantitative results used in the thesis, comparing classical PID guidance, single-agent RL, and the proposed zero-sum MARL methods under combined uncertainty scenarios.

Table 1: Performance comparison under combined uncertainty scenarios. Values are mean ± standard deviation over 1000 test episodes.

Algorithm	Trajectory error (m)	Fuel (m/s)	Success rate (%)	Robustness score (1–5)
PID Control	8432 ± 2156	45.2 ± 8.3	72.4	2
DDPG	1234 ± 892	28.7 ± 5.2	84.6	3
TD3	967 ± 654	26.4 ± 4.1	88.2	4
SAC	1045 ± 721	27.8 ± 4.8	86.9	4
PPO	1398 ± 978	31.2 ± 6.3	81.5	3
MA–DDPG	892 ± 423	25.1 ± 3.2	91.7	4
MA–TD3	687 ± 312	23.4 ± 2.8	95.3	5
MA–SAC	734 ± 367	24.2 ± 3.1	93.8	5
MA–PPO	856 ± 445	26.7 ± 3.9	90.4	4

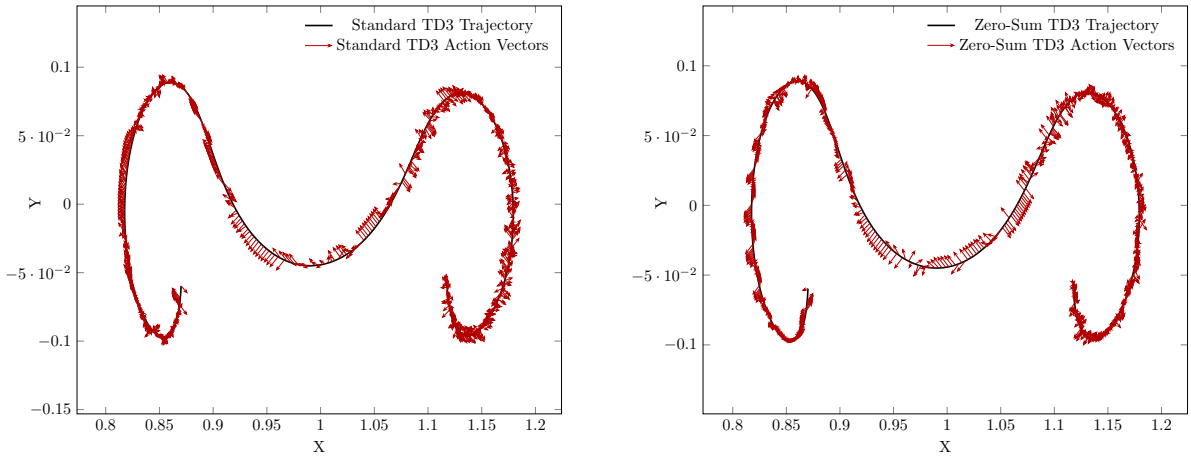
3 Representative Trajectory Tracking Results

Figure 1 compares the nominal TD3 controller with the zero-sum MA–TD3 variant on a representative Earth–Moon CR3BP transfer. The top row shows the tracked trajectories, while the bottom row overlays the applied low-thrust control vectors.



(a) Standard TD3 trajectory.

(b) Zero-sum MA–TD3 trajectory.



(c) Standard TD3 with control vectors.

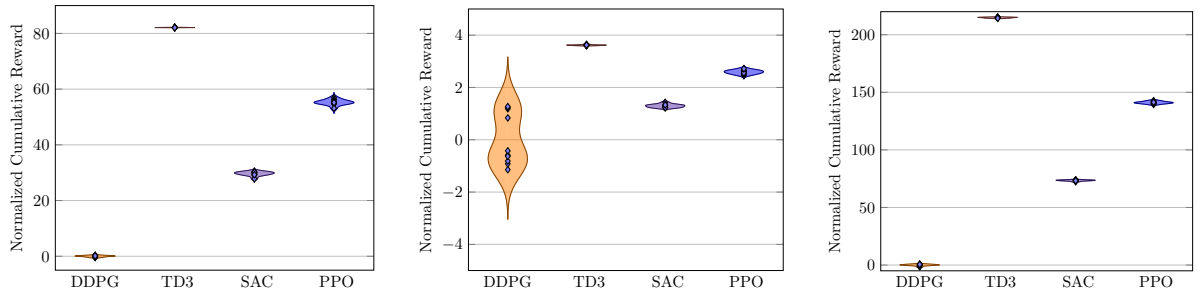
(d) Zero-sum MA–TD3 with control vectors.

Figure 1: Trajectory tracking performance of standard TD3 vs. zero-sum MA–TD3 in the Earth–Moon CR3BP. The zero-sum controller achieves tighter tracking with smoother and more fuel-efficient thrust profiles.

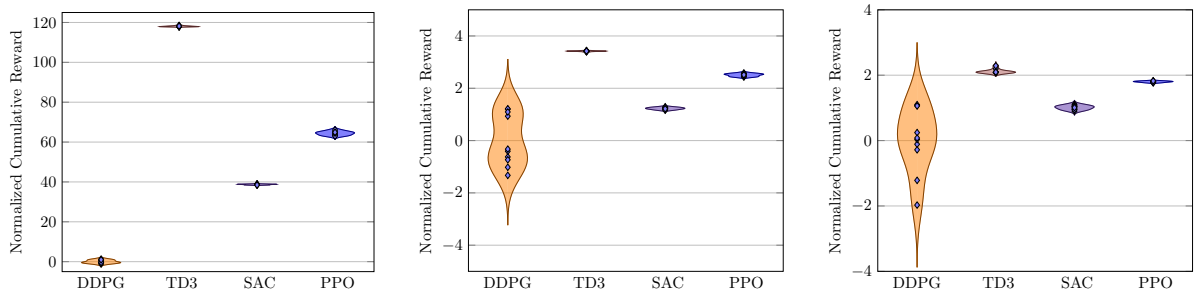
4 Robustness Analysis Under Uncertainty

4.1 Zero-Sum Multi-Agent RL (All Algorithms Combined)

Figure 2 shows the distribution of normalized cumulative rewards for the four algorithms (DDPG, TD3, SAC, PPO) under six different uncertainty scenarios when trained in the proposed zero-sum two-player setting.



(a) Actuator disturbance. (b) Sensor noise. (c) Initial condition shift.

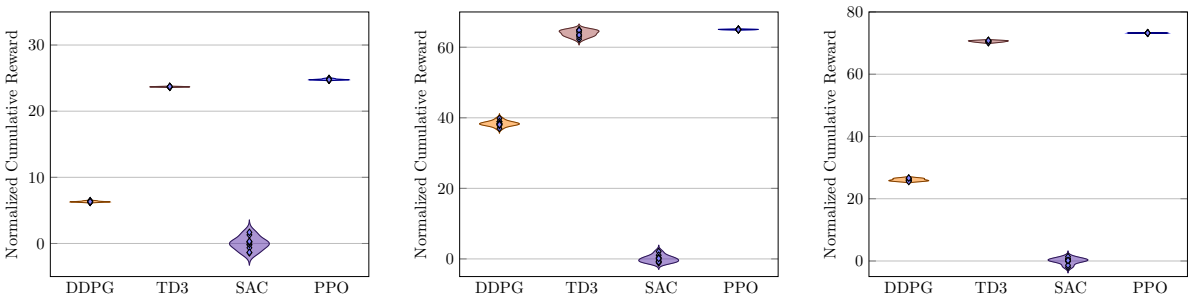


(d) Time delay. (e) Model mismatch. (f) Partial observation.

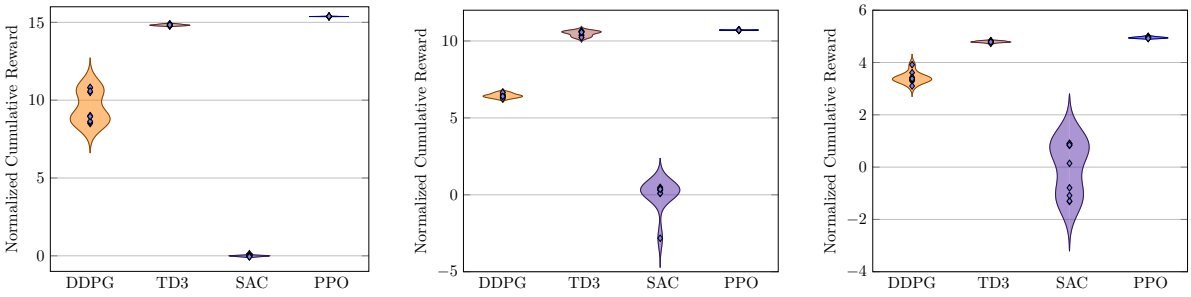
Figure 2: Zero-sum multi-agent RL performance across six uncertainty scenarios. Each violin shows the distribution of normalized cumulative reward over multiple test trajectories.

4.2 Standard Single-Agent RL (All Algorithms Combined)

For comparison, Figure 3 reports the same evaluation for standard single-agent RL training of DDPG, TD3, SAC, and PPO.



(a) Actuator disturbance. (b) Sensor noise. (c) Initial condition shift.



(d) Time delay. (e) Model mismatch. (f) Partial observation.

Figure 3: Standard single-agent RL performance under the same uncertainty scenarios as Figure 2. The zero-sum formulation leads to higher median performance and tighter distributions.