Robust Reinforcement Learning Differential Game Guidance in Low-Thrust, Multi-Body Dynamical Environments

A Zero-Sum Reinforcement Learning Approach in Three-Body Dynamics

Ali Baniasad

Supervisor: Dr. Nobahari

Department of Aerospace Engineering
Sharif University of Technology

Outline

Research Motivation

- **Space missions** increasingly require autonomous guidance systems
- **Low-thrust spacecraft** operate in complex gravitational environments
- **Three-body dynamics** (Earth-Moon CRTBP) present inherent instabilities
- **Classical control methods** struggle with:
  - Model uncertainties
  - Environmental disturbances
  - Fuel efficiency requirements
- **Need for robust, adaptive guidance** without precise dynamic models

How can we achieve robust spacecraft guidance in uncertain environments?

Problem Statement

Research Objective

Design a robust guidance framework for low-thrust spacecraft operating in Earth-Moon three-body dynamics under uncertainties.

System Characteristics:

- State: $\mathbf{x} = [x, y, \dot{x}, \dot{y}]^T$
- Control: $\mathbf{u} \leq u_{\max}$
- Dynamics: $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$

Mission Environment:

- Earth-Moon CRTBP
- Lyapunov orbit transfer
- Low-thrust propulsion

Mathematical Formulation

Optimal control problem with state-dependent uncertainties and adversarial disturbances

Problem Statement - Challenges & Approach

Uncertainty Sources:

- Random initial conditions
- Actuator disturbances
- Sensor noise
- Model mismatch
- Communication delays
- External perturbations

Design Requirements:

- Robust performance
- Fuel efficiency
- Real-time capability
- Model-free approach
- Adaptive behavior
- Safety guarantees

Key Challenge

Formulate as zero-sum differential game: spacecraft vs. environmental disturbances

Multi-Agent Reinforcement Learning Framework

Game-Theoretic Formulation:

- **Player 1**: Controller agent (spacecraft)
- **Player 2**: Adversary agent (disturbances)
- **Objective**: Zero-sum differential game

Training Paradigm:

- Centralized training
- Decentralized execution

Algorithm Extensions

Single-Agent → Multi-Agent Zero-Sum

Extended four continuous RL algorithms to handle adversarial scenarios

DDPG → MA-DDPG

- Deterministic policy gradient
- Experience replay
- Target networks

SAC → MA-SAC

- Entropy regularization
- Stochastic policies
- Automatic temperature tuning

TD3 → MA-TD3

- Twin critic networks
- Delayed policy updates
- Target policy smoothing

PPO → MA-PPO

- Proximal policy optimization
- Clipped surrogate objective
- On-policy learning

Network Architecture

Actor-Critic Structure:

- 3 hidden layers
- 32 neurons per layer
- ReLU activation
- Learning rate: 3×10^{-4}

Training Parameters:

- 1M environment steps
- Batch size: 1024
- Buffer size: 100K
- Adam optimizer

Evaluation Scenarios

Mission: Transfer to Earth-Moon Lyapunov Orbit

Circular Restricted Three-Body Problem (CRTBP) environment

Uncertainty Scenarios:

1. Random initial states ($\sigma = 0.1$)
2. Actuator disturbances ($\sigma = 0.05$)
3. Model mismatch ($\sigma = 0.05$)
4. Partial observations (50% missing)
5. Sensor noise ($\sigma = 0.05$)
6. Communication delays (10 steps)

Performance Metrics:

- Trajectory accuracy
- Fuel consumption
- System stability
- Convergence rate
- Robustness measures

Baseline Comparison:

- Single-agent variants

Trajectory Comparison: MA-TD3 Performance

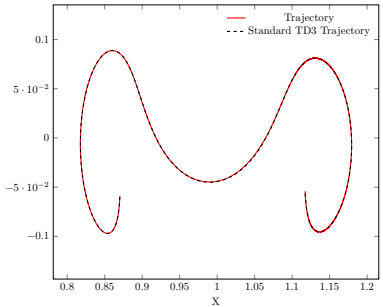


Figure: Single-Agent vs Multi-Agent TD3

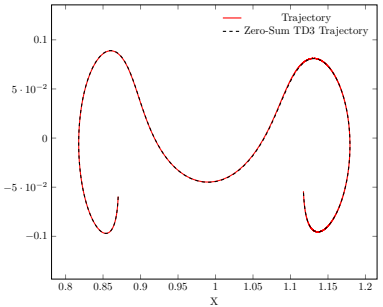


Figure: Thrust Commands

Key Observation

Zero-sum variants demonstrate more direct trajectories with smoother thrust profiles and reduced fuel consumption

Robustness Evaluation Results

Algorithm Performance Comparison

Single-Agent Algorithms:

Multi-Agent Zero-Sum:

Real-Time Deployment

Optimization Techniques:

- INT8 quantization
- ONNX format export
- ROS 2 integration
- Hardware-in-the-loop testing

Performance Metrics:

- Inference time: 5.8 ms
- Memory usage: 9.2 MB
- Control frequency: 100 Hz
- Zero deadline misses

Research Contributions

1. **Novel Framework Development**
   - First multi-agent zero-sum RL framework for spacecraft guidance in CRTBP
   - Game-theoretic formulation for robust control under uncertainties
2. **Algorithm Extensions**
   - Extended four major RL algorithms to multi-agent zero-sum variants
   - Comprehensive performance analysis across uncertainty scenarios
3. **Practical Implementation**
   - Real-time deployment with quantization and optimization
   - Hardware-in-the-loop validation on ROS 2 platform
4. **Performance Validation**
   - Demonstrated superior robustness compared to classical methods
   - Quantified improvements in fuel efficiency and trajectory accuracy

Conclusions

Main Findings

- Multi-agent zero-sum RL enables robust spacecraft guidance without precise dynamic models
- MATD3 delivers optimal performance across all evaluation scenarios
- Framework is ready for practical deployment with real-time constraints

Advantages:

- Model-free approach
- Adaptive to uncertainties
- Fuel-efficient trajectories
- Real-time capability

Future Work:

- 3D CRTBP extension
- Multi-body dynamics
- Swarm coordination
- Deep space missions

Thank You!

Questions & Discussion

Ali Baniasad

Department of Aerospace Engineering
Sharif University of Technology

Appendix - Technical Implementation Details

twidthtwidthextbfEnvironment Setup:

- PyTorch & OpenAI Gym
- CRTBP dynamics im-

twidthtwidthextbfTraining Configuration:

- Episodes: 1M steps
- Batch size: 1024

Appendix - Zero-Sum Algorithm Extensions

MATD3 Algorithm Highlights

- Twin critic networks to reduce overestimation bias
- Delayed policy updates for stability
- Target policy smoothing for robustness
- Competitive training between agents

twidthtwidthextbfKey Mod-i-fi-ca-tions:

1. Shared

twidthtwidthextbfPerformance Ben-e-fits:

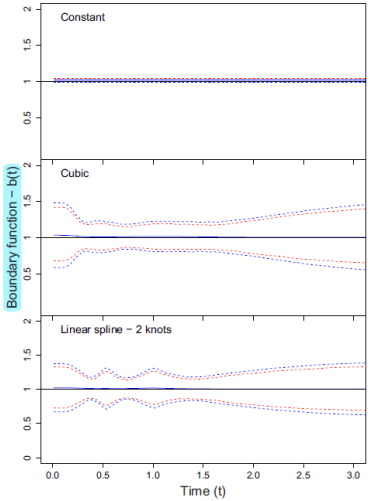
1. Enhanced ro-bust-

Appendix - Future Research Directions

1. extbfMulti-Body Extensions
   - Sun-Earth-Moon system
   - Jupiter-Europa dynamics
   - Asteroid belt navigation
2. extbfSwarm Coordination
   - Formation flying control
   - Distributed decision making
   - Communication constraints
3. extbfAdvanced Techniques
   - Meta-learning approaches
   - Hierarchical reinforcement learning
   - Physics-informed neural networks
4. extbfMission Applications
   - Lunar gateway operations
   - Deep space exploration
   - Interplanetary transfers

Appendix - A figure

Return to presentation



Appendix - Terms

Some Estimators:

- Drift: $\hat{\delta}$
- Boundary: $\hat{b}(t)$

Some Variables:

- $\hat{V}$
- $\hat{m}_S$
- $\bar{m}$
- $m_J(\tau)$

Return to presentation

Appendix - Definitions

1. A definition

Return to presentation

Appendix - Theorems

1. A theorem

Return to presentation