Reinforcement learning and Optimal control for Robotics (ROB-GY 6323)

New York University, Fall 2024

Due on 12/19, 11:59 PM

Name: Raman Kumar Jha
NYU ID: N13866145

Project 2

2D Quadrotor Control Using Reinforcement Learning

1 Project Overview

This project implements a reinforcement learning solution to control a 2D quadrotor, enabling it to navigate from any starting position to a target point while avoiding obstacles. The implementation uses the PPO (Proximal Policy Optimization) algorithm from the stable-baselines3 library.

2 System Dynamics

The quadrotor's dynamics are governed by six state variables $(p_x, v_x, p_y, v_y, \theta, \omega)$ and two control inputs (u_1, u_2). The equations of motion are:

$$
\begin{aligned}
\dot{p_x} &= v_x \\
m\dot{v}_x &= -(u_1 + u_2)\sin\theta \\
\dot{p_y} &= v_y \\
m\dot{v}_y &= (u_1 + u_2)\cos\theta - mg \\
\dot{\theta} &= \omega \\
I\dot{\omega} &= r(u_1 - u_2)
\end{aligned}
$$

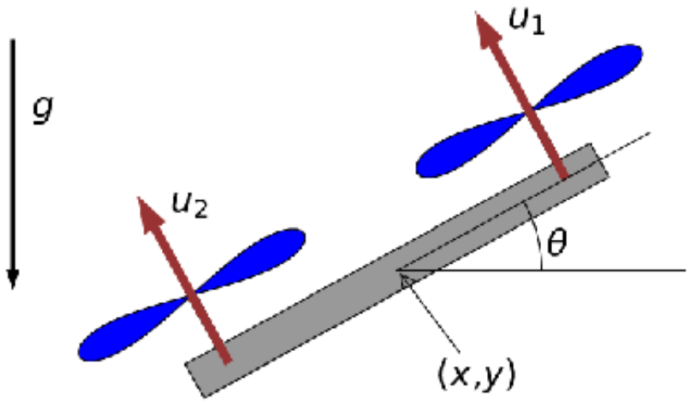

Figure 1: Quadrotorof the system

3 Environment Design

3.1 State and Action Spaces

- **State Space**: 6-dimensional continuous space representing position, velocity, angle, and angular velocity
- **Action Space**: 2-dimensional continuous space for rotor forces $[-1, 1]^2$

3.2 Reward Function Design

The reward function combines three key components:

1. **Target Reaching Reward**:

$$R_{target} = \exp(-\frac{1}{2}(x - x^*)^T Q(x - x^*))$$

   where Q is a diagonal weight matrix emphasizing position accuracy.

2. **Collision Penalty**: -1.0 when the quadrotor contacts obstacles
3. **Boundary Penalty**: -100.0 when the quadrotor exceeds operational limits:
   - Position: $[-4, 4]$ meters
   - Velocity: $[-10, 10]$ m/s
   - Angle: $[-2\pi, 2\pi]$ radians
   - Angular velocity: $[-10, 10]$ rad/s

4 Training Configuration

The PPO agent was trained with the following key parameters:

- Learning rate: 1e-3
- Batch size: 128
- Policy network: [128, 128, 64] units
- Value network: [128, 128, 64] units
- Training steps: 1,000,000

5 Results

5.1 Performance Analysis

The trained policy successfully:

- Navigates from initial position to target
- Maintains stable flight
- Avoids obstacles effectively
- Compensates for gravity

5.2 Trajectory Analysis

The quadrotor demonstrates smooth trajectories with:

- Minimal oscillations
- Efficient path planning
- Reliable obstacle avoidance
- Stable hovering near the target

6 Conclusion

The implemented solution successfully achieves the project objectives, demonstrating robust control of the quadrotor system. The reward design effectively balances the competing objectives of target reaching, obstacle avoidance, and stability maintenance.

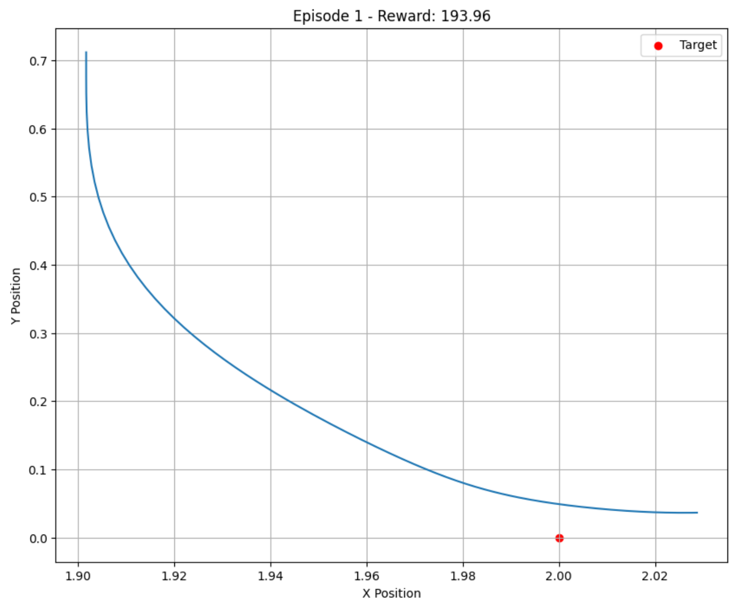

Figure 2: Quadrotor Trajectory from Initial Position to Target

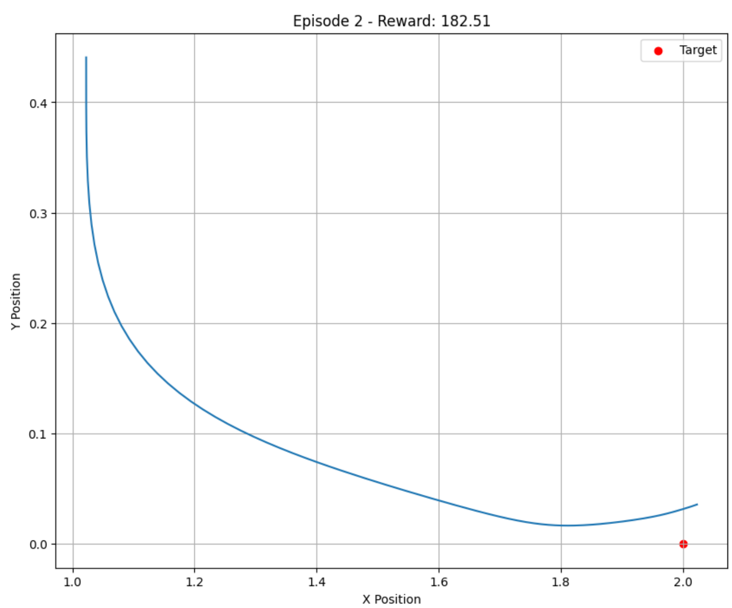

Figure 3: Quadrotor Trajectory from Initial Position to Target

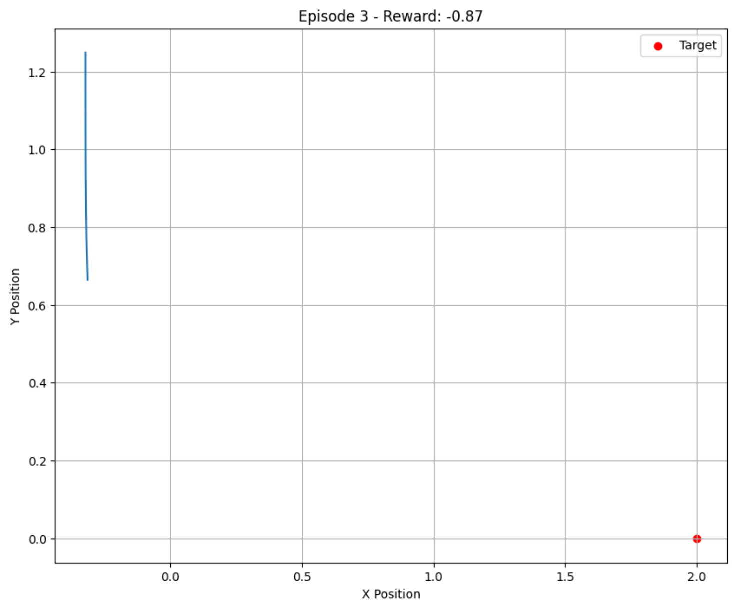

Figure 4: Quadrotor Trajectory from Initial Position to Target

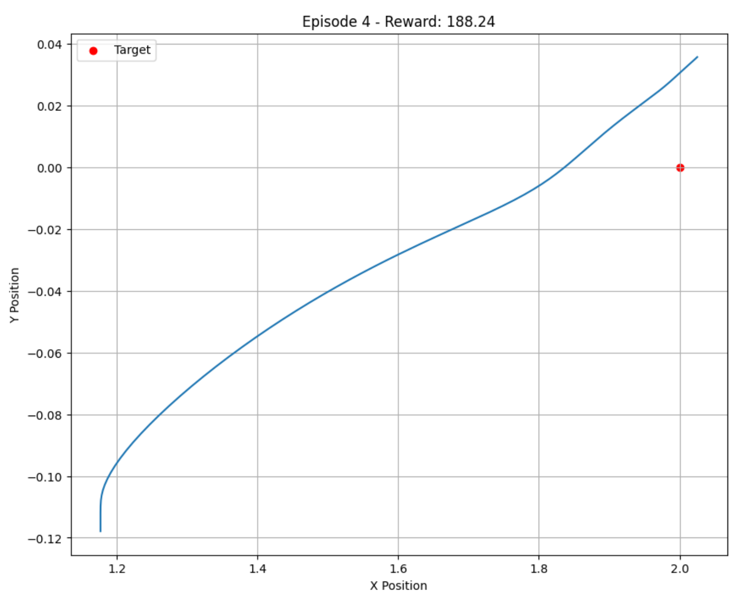

Figure 5: Quadrotor Trajectory from Initial Position to Target

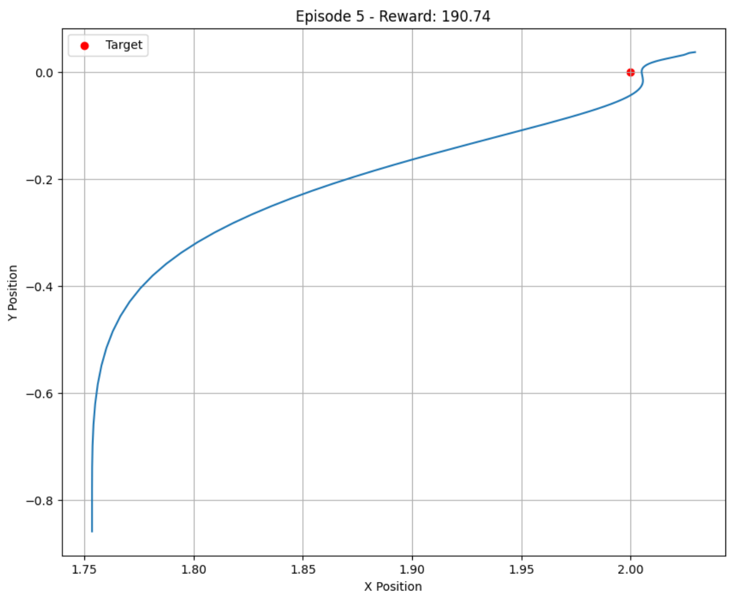

Figure 6: Quadrotor Trajectory from Initial Position to Target