Reinforcement Learning and Optimal Control Project 1

Rishabh Verma

December 2024

1 Problem Statement

The goal of this project is to control a 2D quadrotor to reach a target while avoiding obstacles using a learned policy. We create a custom environment using stable_baseline_3 and train the agent using PPO.

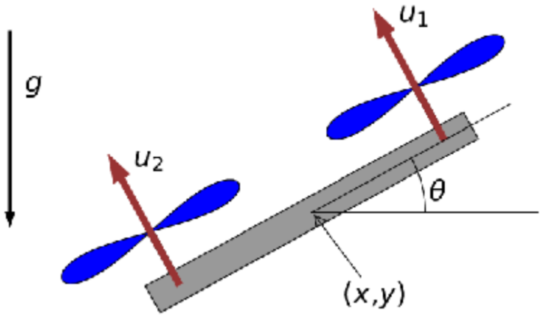

Figure 1: 2D Quadrotor Model

2 Model

2.1 Reward Function

The reward function is made out of three terms:

1. A positive reward to make the quadrotor reach the target given by
$$\exp(-\frac{1}{2}(x - x^\star)Q(x - x^\star) - \frac{1}{2}(u - u_{\text{gravity}})R(u - u_{\text{gravity}})) \tag{1}$$
2. A big penalty of -100 is imposed on the agent if it cross the bounds $p_x \in [-4, 4], \quad v_x \in [-10, 10], \quad p_y \in [-4, 4], \quad v_y \in [-10, 10], \quad \theta \in [-2\pi, 2\pi], \quad \omega \in [-10, 10]$
3. A penalty of -1 is imposed if collision is detected on the robot.

The step function follows the algorithm below:

Algorithm 1 Quadrotor Reward Calculation

```
Call next state of quadrotor
Obtain the reward from equation 1
if state out of bounds then
    Terminate and reward ← reward - 100
else
    if collision then
        reward ← reward - 1
    end if
    Increase the step
    Check if the step reached horizon
end if
```

3 Results

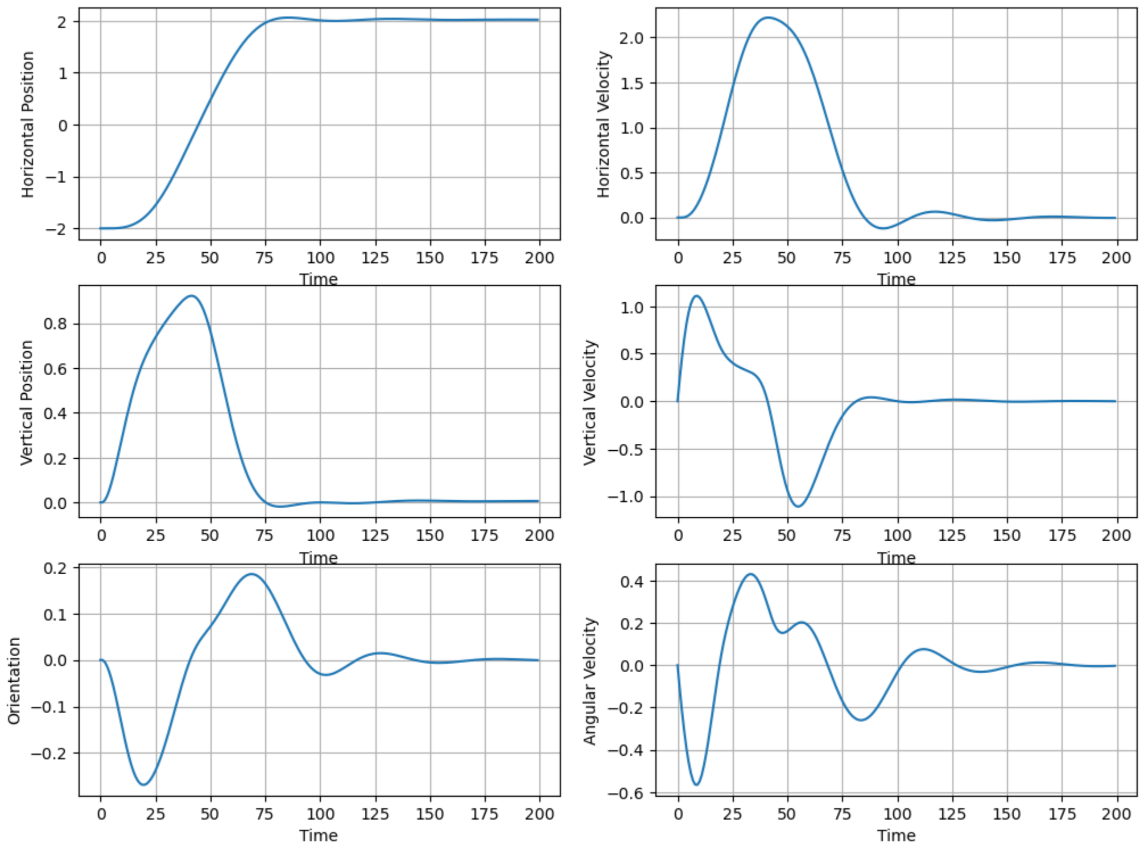

Figure 2: Quadrotor States

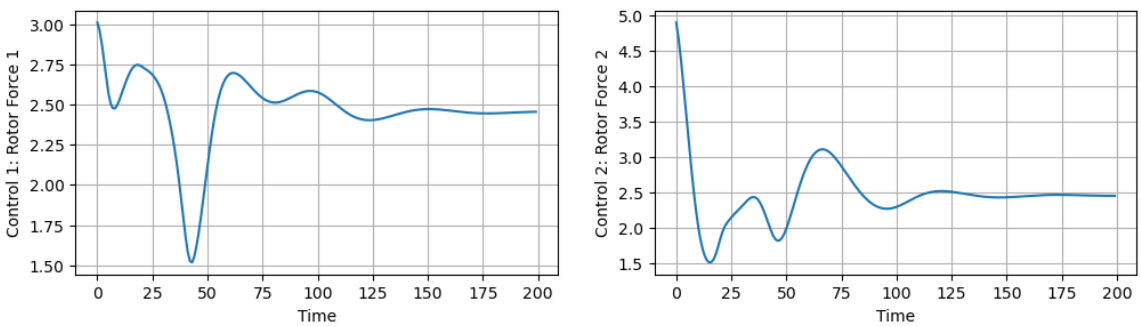

Figure 3: Control Inputs