Reinforcement Learning and Optimal Control for Robotics-Project-2-Report

vp2504

December 2024

Reward Design for Quadrotor Environment

Key Components of the Reward Function

1. **State Error:** Penalizes deviation from target state:

$$\text{state\_cost} = \frac{1}{2}(\mathbf{s} - \mathbf{s}^*)^T \mathbf{Q}(\mathbf{s} - \mathbf{s}^*).$$

2. **Control Effort:** Penalizes deviation of action from baseline:

$$\text{control\_cost} = \frac{1}{2}(\mathbf{a} - \mathbf{u}_{\text{gravity}})^T \mathbf{R}(\mathbf{a} - \mathbf{u}_{\text{gravity}}).$$

3. **Goal-Oriented Reward:**

$$r_{\text{goal}} = \exp\left(-\text{state\_cost} - \text{control\_cost}\right).$$

4. **Collision Penalty:**

$$r_{\text{collision}} = \begin{cases} -3.0 & \text{if collision,} \\ 0.0 & \text{otherwise.} \end{cases}$$

5. **Out-of-Bounds Penalty:**

$$r_{\text{out\_of\_bounds}} = \begin{cases} -100.0 & \text{if out of bounds,} \\ 0.0 & \text{otherwise.} \end{cases}$$

Combined Reward Function

$$\text{reward} = r_{\text{goal}} + r_{\text{collision}} + r_{\text{out\_of\_bounds}}.$$

Termination Conditions

- Terminate if out of bounds
- Truncate if steps exceed max_steps.

Weight Matrices

$$\mathbf{Q} = \text{diag}(1, 0.1, 1, 0.1, 0.1, 0.01), \quad \mathbf{R} = 0.01 \cdot \mathbf{I}_2.$$

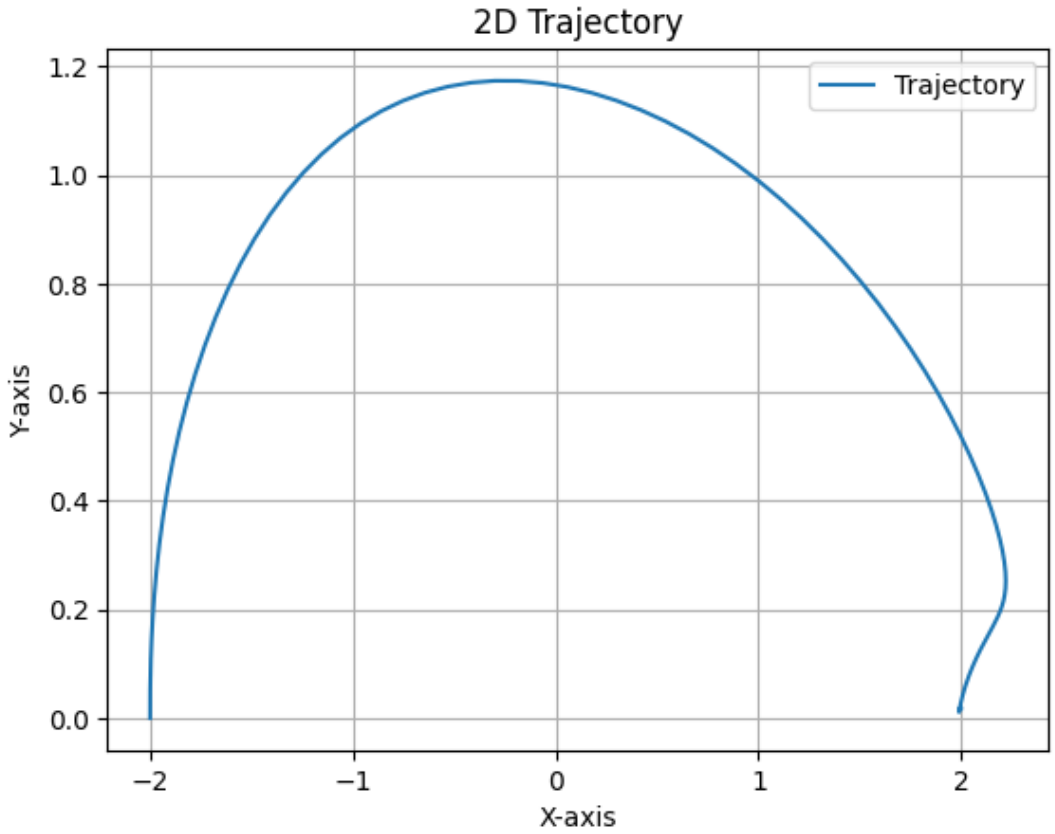

Figure 1: Trajectory of a Quadrotor

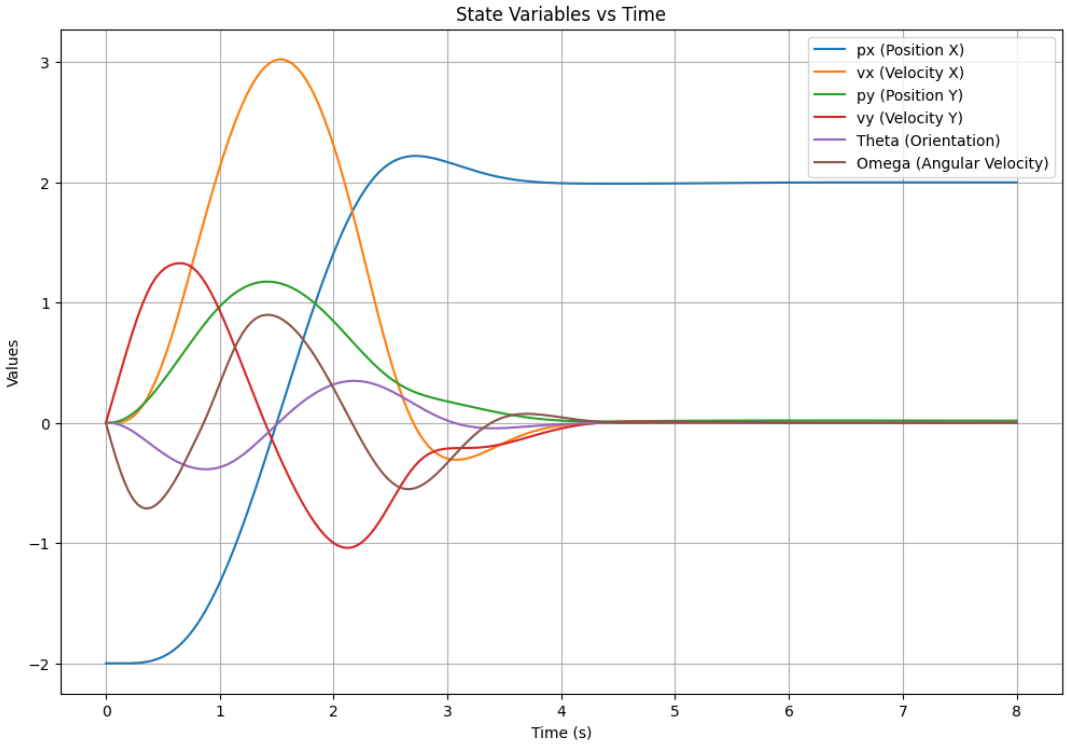

Figure 2: States of a Quadrotor w.r.t time