Exploring the Dynamics of Reinforcement Learning in Aerospace Control

Georg Schäfer

Josef Ressel Center for Intelligent and Secure Industrial Automation
Department for Information Technologies and Digitalisation
Salzburg University of Applied Sciences

February 7, 2024

Salzburg University
of Applied Sciences

Topics of Interest

JRC ISIA

- Autonomous agents
- Process optimization

Topics of Interest

JRC ISIA

- Autonomous agents
- Process optimization

AI4Green

- Process optimization
- Energy efficiency and cost reduction

Topics of Interest

JRC ISIA

- Autonomous agents
- Process optimization

AI4Green

- Process optimization
- Energy efficiency and cost reduction

What test system?

Requirements

- Mechatronic system
- Multiple inputs / Multiple outputs
- Non-linear system
- System model can be derived

System Selection

Requirements

- Mechatronic system
- Multiple inputs / Multiple outputs
- Non-linear system
- System model can be derived

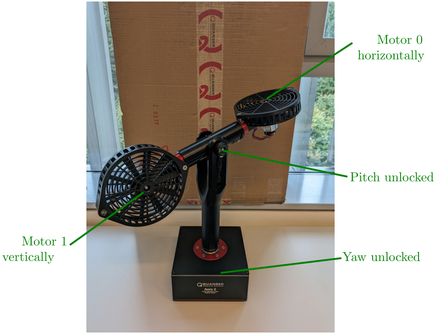


Figure: Aero 2 system (`https://quanser.com`)

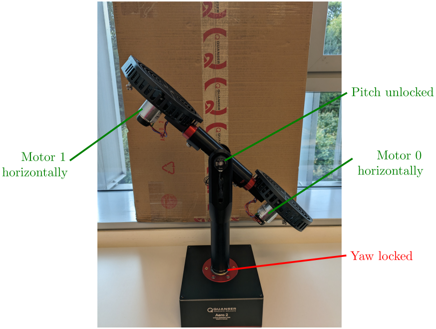


Figure: 1 degree-of-freedom (DOF) - Pitch control (one input $u = u_0 = -u_1$, one output $y = \Theta$)

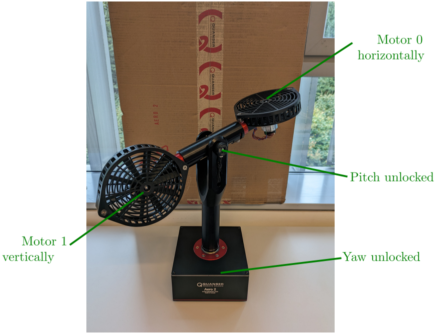


Figure: 2 DOF - Yaw and Pitch control (two inputs, two outputs)

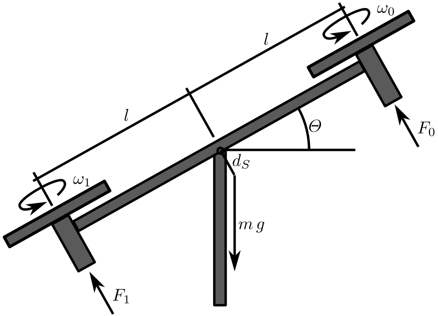


Figure: 1 DOF - Pitch control (one input $u = u_0 = -u_1$, one output $y = \Theta$)

- 1 DOF configuration, controlled variable y: pitch-angle, manipulated variable u: fan voltage ($u_0 = -u_1 = u$)
- Thrusters: $F_i \approx k \cdot u_i$
- Beam:

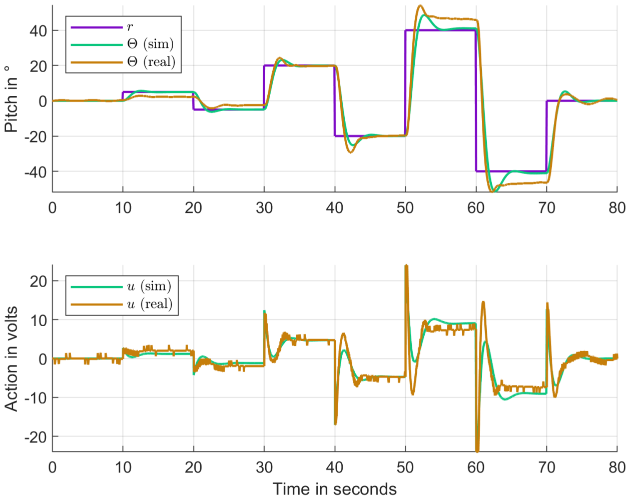
$$\frac{d\Theta}{dt} = \omega$$

$$J_p \frac{d\omega}{dt} = \underbrace{(F_0 - F_1)\, l}_{2\,k\,u\,l} - D_p \omega - m\, g\, d_S \sin(\Theta)$$

Model Predictive Control

- works well in simulation
- steady state deviation on real system, depending on angle

Can we obtain similar results via Reinforcement Learning?



Reinforcement Learning Setting

Problem Formulation

Orient the beam to a desired angle (r) by applying a voltage (u and $-u$) to the motors.

Reinforcement Learning Setting

Problem Formulation

Orient the beam to a desired angle (r) by applying a voltage (u and $-u$) to the motors.

- **Action**: Voltage applied to the motors (u)
- **State**: Distance to the desired angle ($\Delta = \Theta - r$) and the current angular velocity ($\omega = \dot{\Theta}$)
- **Reward**: Negative absolute distance to the desired angle ($-|\Delta|$)

Reinforcement Learning Setting

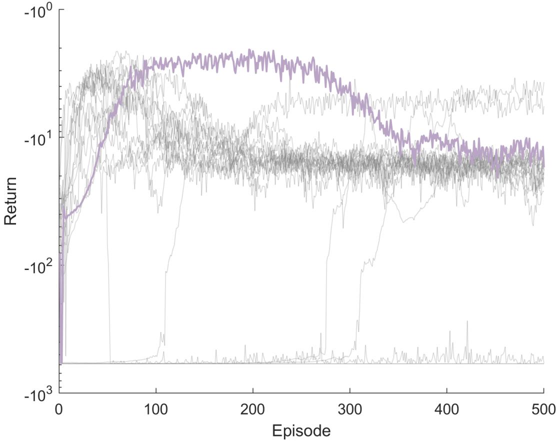
Problem Formulation

Orient the beam to a desired angle (r) by applying a voltage (u and $-u$) to the motors.

- **Action**: Voltage applied to the motors (u)
- **State**: Distance to the desired angle ($\Delta = \Theta - r$) and the current angular velocity ($\omega = \dot{\Theta}$)
- **Reward**: Negative absolute distance to the desired angle ($-|\Delta|$)

- **Agent**: Proximal Policy Optimization (PPO)

RL Training Runs (constant r)



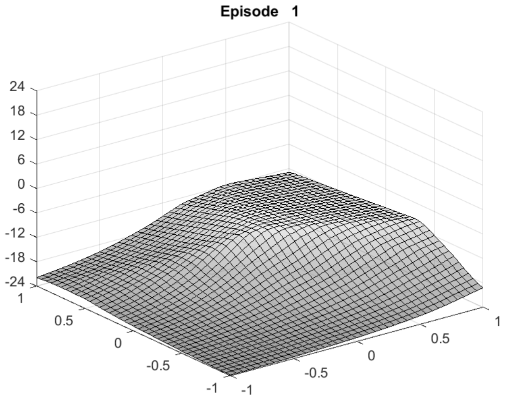
Configurations:

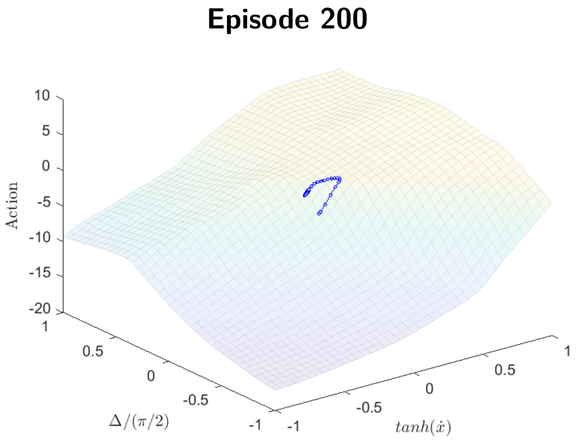
- 20 runs
- $r = 10°$
- Episode length $=$ 60s
- Sample time $=$ 0.1s

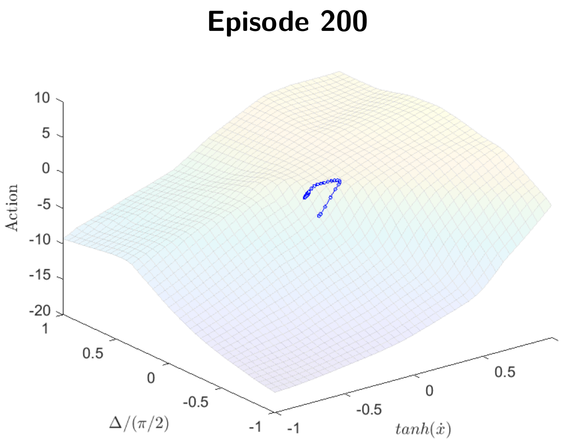
Result:

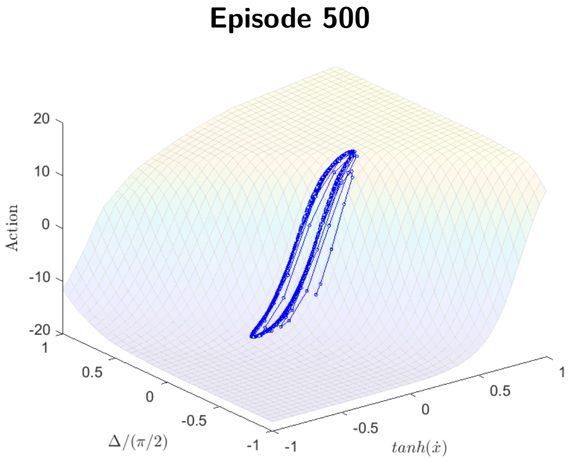
- max(Return) $=$ -2.13
- $\frac{|-2.13|}{600} \cdot \frac{180}{\pi} = 0.64°$

Policy Evaluation

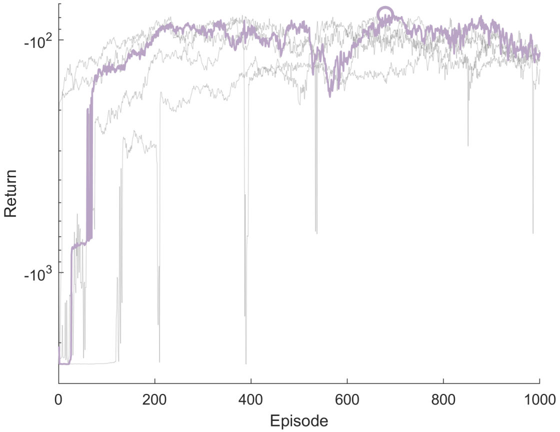



Episode 200
10
5
0
-5
-10
-15
-20
Action
1
0.5
0
-0.5
-1
$\Delta/(\pi/2)$
-1
-0.5
0
0.5
1
$tanh(\dot{x})$


Episode 200
10
5
0
-5
-10
-15
-20
Action
1
0.5
0
-0.5
-1
$\Delta/(\pi/2)$
-1
-0.5
0
0.5
1
$tanh(\dot{x})$


Episode 500
20
10
0
-10
-20
Action
1
0.5
0
-0.5
-1
$\Delta/(\pi/2)$
-1
-0.5
0
0.5
1
$tanh(\dot{x})$

RL Training Runs (dynamic r)

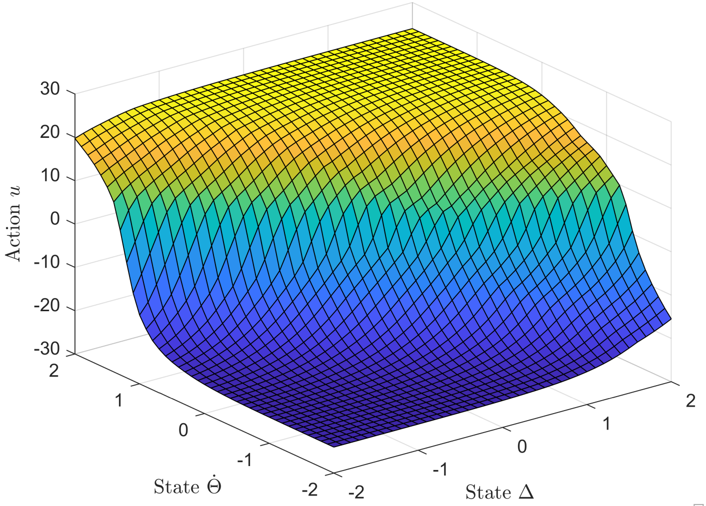


Configurations:

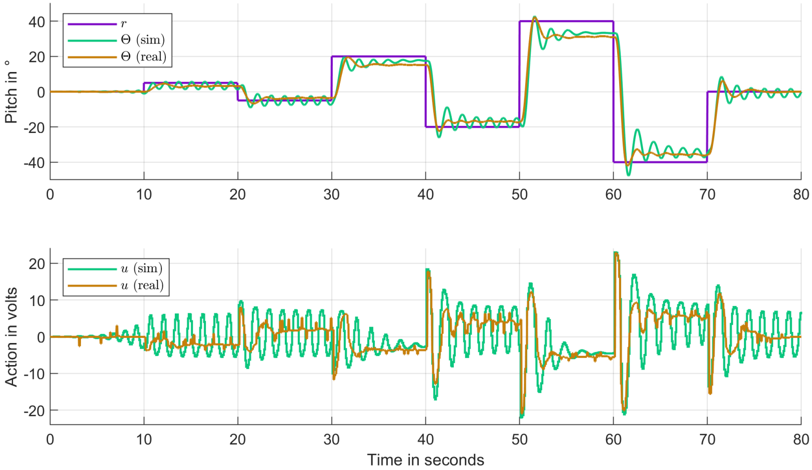
- 5 runs
- $r = [0, -5, 5, 20, -20, 40, -40]$
- r changes every 10s
- Episode length $=$ 80s
- Sample time $=$ 0.1s

Result:

- max(Return) $=$ -77.93
- $\frac{|-77.93|}{800} \cdot \frac{180}{\pi} = 5.6^\circ$


30
20
10
0
-10
-20
-30
Action u
2
1
0
-1
-2
State $\dot{\Theta}$
-2
-1
0
1
2
State Δ

Test Runs



r
Θ (sim)
Θ (real)
Pitch in °
40
20
0
-20
-40
0
10
20
30
40
50
60
70
80
u (sim)
u (real)
Action in volts
20
10
0
-10
-20
Time in seconds

Conclusion and Outlook

Results

Using RL we can achieve a similar performance as with the MPC controller.

Outlook

- Detailed comparison to well established controller like MPC
- Adaptations in state space
- Extension to 2 DOF system