Department of Instrumentation
and Control Engineering

Reinforcement Learning Framework for Adaptive Control of Quadrotor UAVs

Summer Internship presentation

Name: Shreehan S Kate
Roll number: 110122101
shreehan1912@gmail.com

Reinforcement Learning Framework for Adaptive Control of Quadrotor UAVs

Abstract

Nonlinearity, underactuated configuration, and vulnerability to uncertainties and disturbances make the Quadrotor Unmanned Aerial Vehicles (UAVs) a complex control problem. Feedback linearization is a nonlinear control technique that leverages nonlinear transformations to obtain a linearized model of the original system. This transformation facilitates the application of linear control strategies. Fixed-gain PID controllers often suffer from performance degradation caused by parameter variations and external disturbances. This study proposes an adaptive PID controller, leveraging the Deep Reinforcement Learning (DRL) framework for altitude and attitude control. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm is employed to tune the PID controller gains in an online, model-free environment. The TD3 agent learns control policies that dynamically modify PID gains in response to observed system behavior through continuous interaction with the environment. The performance of the TD3-driven PID controller is assessed through extensive simulations performed on the feedback-linearized quadrotor model under varying altitude trajectories and external disturbances, followed by a comparative analysis between the proposed controller, fixed-gain PID and other commonly known linear control strategies to evaluate improvements in robustness and tracking.

Areas of Research

This project focuses on advanced control strategies, specifically reinforcement learning-based adaptive PID controllers to enhance altitude tracking, stability, and adaptability under dynamic conditions.

Linear Control Theory

This project explores linear control algorithms such as Proportional-Integral-Derivative Control (PID), Linear Quadratic Regulator (LQR), and Linear Model Predictive Control (LMPC), designed using simplified dynamic models near steady-state flight, as a foundation for developing more advanced and adaptive control strategies

Non-linear Control Theory

Nonlinear control is explored through feedback linearization, enabling exact cancellation of system nonlinearities to achieve improved tracking performance across a wider range of operating conditions.

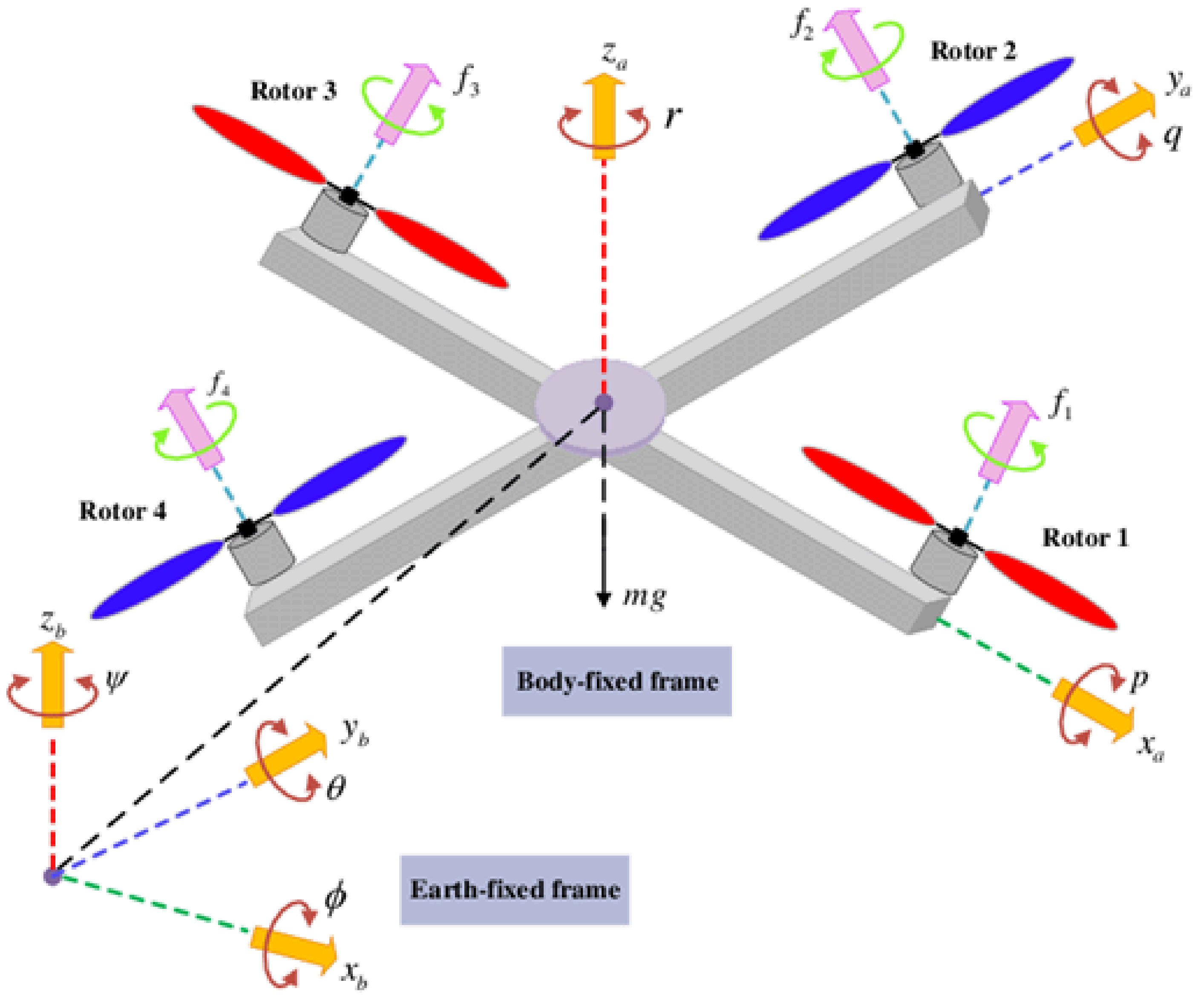
Reinforcement Learning

Reinforcement learning is employed to develop adaptive, data-driven control policies that can learn optimal actions through interaction with the quadrotor environment.

Relevance of Aerial Robotics

- Aerial robotics research has grown to **automate** labor-intensive tasks.
- Widely used in agriculture, logistics, surveillance, and **environmental monitoring**.
- Optimized for tasks like package delivery, surveillance, and crop monitoring.
- Military use for **reconnaissance** has driven technological advances.
- Rising interest in environmental uses like **pollution detection** and **wildlife tracking**.
- Supports disaster response with real-time data, rescue aid, and supply delivery.

Literature Survey



- Conducted a thorough investigation of the literature present that deals with quadrotor dynamics.
- The literature survey helped me understand that the Quadrotor UAV is a complex control problem that is actively researched due to its underactuated nature. A typical quadrotor UAV has 6 state variables or DOF and 4 control inputs.
- Understood the several nonlinear dynamical equations proposed in the literature.

Application of Nonlinear Control

- Once I got a good grasp of the nonlinear UAV dynamics, I moved on towards studying nonlinear control techniques in depth.

- Some of the nonlinear control techniques are as follows: feedback linearization, Sliding Mode Control, Backstepping control, Lyapunov-based control, and many more.

- Leveraged the feedback linearization technique and derived inverse dynamical equations that would be used to cancel out nonlinearities present in the nonlinear UAV model, and subsequently transformed the system into a linearized model.

- The above formulation yielded four virtual control inputs, which are the inputs to the feedback-linearized quadrotor UAV model.

Linear Control Algorithms

- After implementing the feedback-linearizing equations, the nonlinear system is transformed into a linear model that can be represented in a state-space form.
- A decoupled linear system is obtained, consisting of four virtual control inputs.
- Conducted a literature survey on the various linear control architectures developed for the nonlinear control of UAVs.
- Implementation of three linear control strategies is carried out: PID, LQR, and MPC

PID Controller design

- Designed and tuned four PID controllers responsible for the control of the four decoupled axes (roll, pitch, yaw, and altitude).

- Achieved stable tracking of all four control axes.

- I made sure to incorporate saturation limits on the controller outputs of the four PID controllers in order to avoid actuator damage.

LQR Controller Design

- Solved an optimal control problem, leveraging Linear Quadratic Regulator (LQR) design to achieve full-state feedback control of the feedback-linearized quadrotor UAV

- Implemented the Genetic Algorithm as a metaheuristic optimization technique to select optimal Q and R matrices.

- Computed a 4x8 state feedback gain matrix K and formulated an optimal control law u = -K*(ref-output).

- Achieved stable tracking of all four control variables.

Linear Model Predictive Control

- Linear Model Predictive Control Algorithm is an optimal control technique leveraged to facilitate Multi-Input-Multi-Output Control, making it suitable for a control system such as a quadrotor UAV.

- Formulated the MPC optimization problem to minimize a finite-horizon cost function while satisfying input and state constraints in real time.

- Achieved stable and precise tracking of all four control variables, demonstrating effective constraint handling.

The Deep Reinforcement Learning Approach

- After developing the fixed-gain control architectures, I shifted towards conducting a literature review on reinforcement learning for control.

- Understood the fundamentals of actor-critic deep neural networks and reward function design. Implemented the Twin-Delayed Deep Deterministic Policy Gradient algorithm for the system.

- Designed the actor and the critic neural network for the Reinforcement Learning Agent.

- Implemented dual critics for the TD3 agent to address overestimation bias, commonly known to have occurred in the training of DDPG agents.

Actor-Critic Architecture

TABLE II: Actor Network architecture

Layer	Nodes	Activation Function
Input Layer	16	N/A
First Hidden Layer	256	ReLU + LayerNorm
Second Hidden Layer	256	ReLU + LayerNorm
Output Layer	12	Tanh + Scaling

Actor Network

TABLE III: Critic Network architecture (Per Q-Function)

Layer	Nodes	Activation Function
State Input Layer	16	N/A
Action Input Layer	12	N/A
Hidden Layer (State)	128	ReLU + LayerNorm
Hidden Layer (Action)	128	ReLU + LayerNorm
Merged Layer	128	ReLU
Output Layer	1	Identity

Critic Network

Building the Agent and Reward Function

- Defined the hyperparameters required for training the RL agent. Devised a 16x1 vector representing the states or observations and a 12x1 vector representing the PID gains of the four PID controllers

- Derived a reward function inspired by the Linear Quadratic Gaussian Cost function.

- Leveraged the early stopping criterion technique to promote stable learning.

- Allotted negative penalty for extremely poor tracking and control effort usage and rewarded positively upon low ISE, if observed after the completion of the episode.

$$R = -\sum_{k=0}^{T} \left(w_{\text{error}} \cdot e_k^2 + w_{\text{action}} \cdot u_{k-1}^2 \right)$$

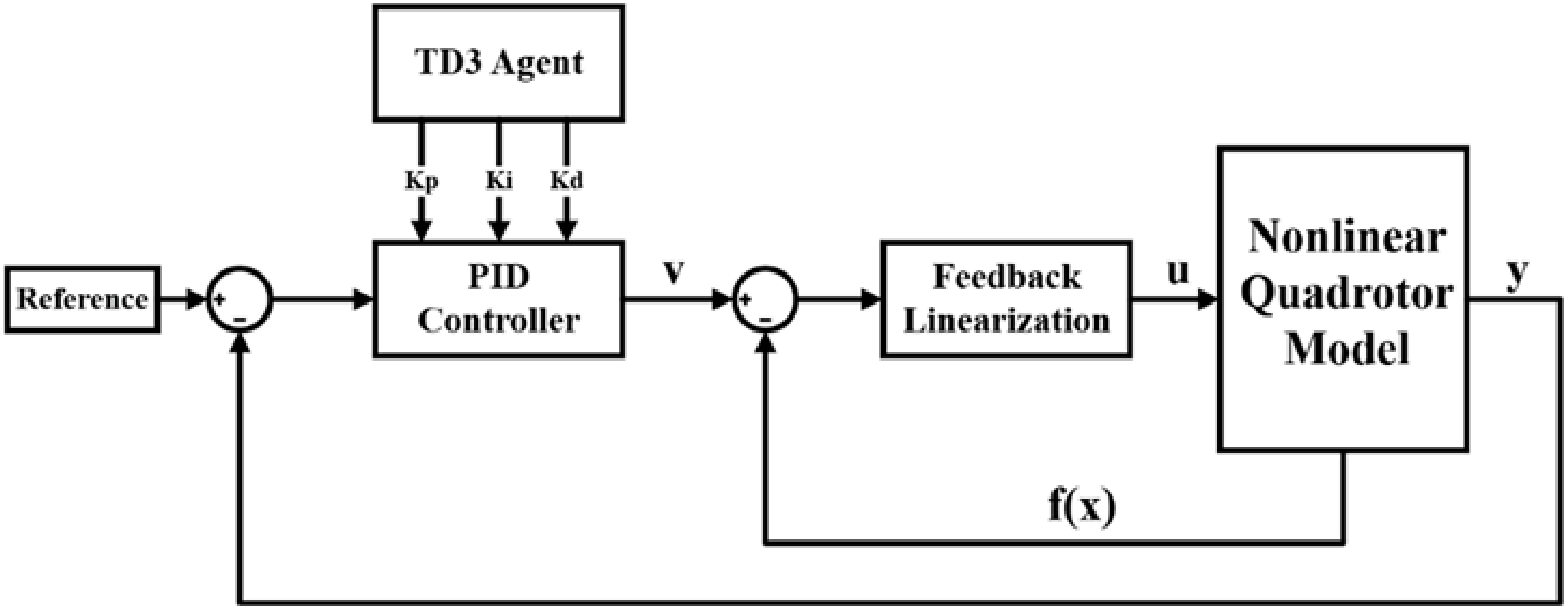
LQG-inspired reward function

- The terms w(error) and w(action) are essentially the weights assigned to each of the four tracking variables. A higher value of weights means large errors and undesirable control usage as a result of sub-optimal control policies will be penalized heavily, resulting in a higher cost.
- The objective of the agent is to converge to an optimal policy, maximizing the cumulative reward.
- The environment was designed in Simulink, and the Reinforcement Learning Agent was developed in MATLAB's Reinforcement Learning Designer App

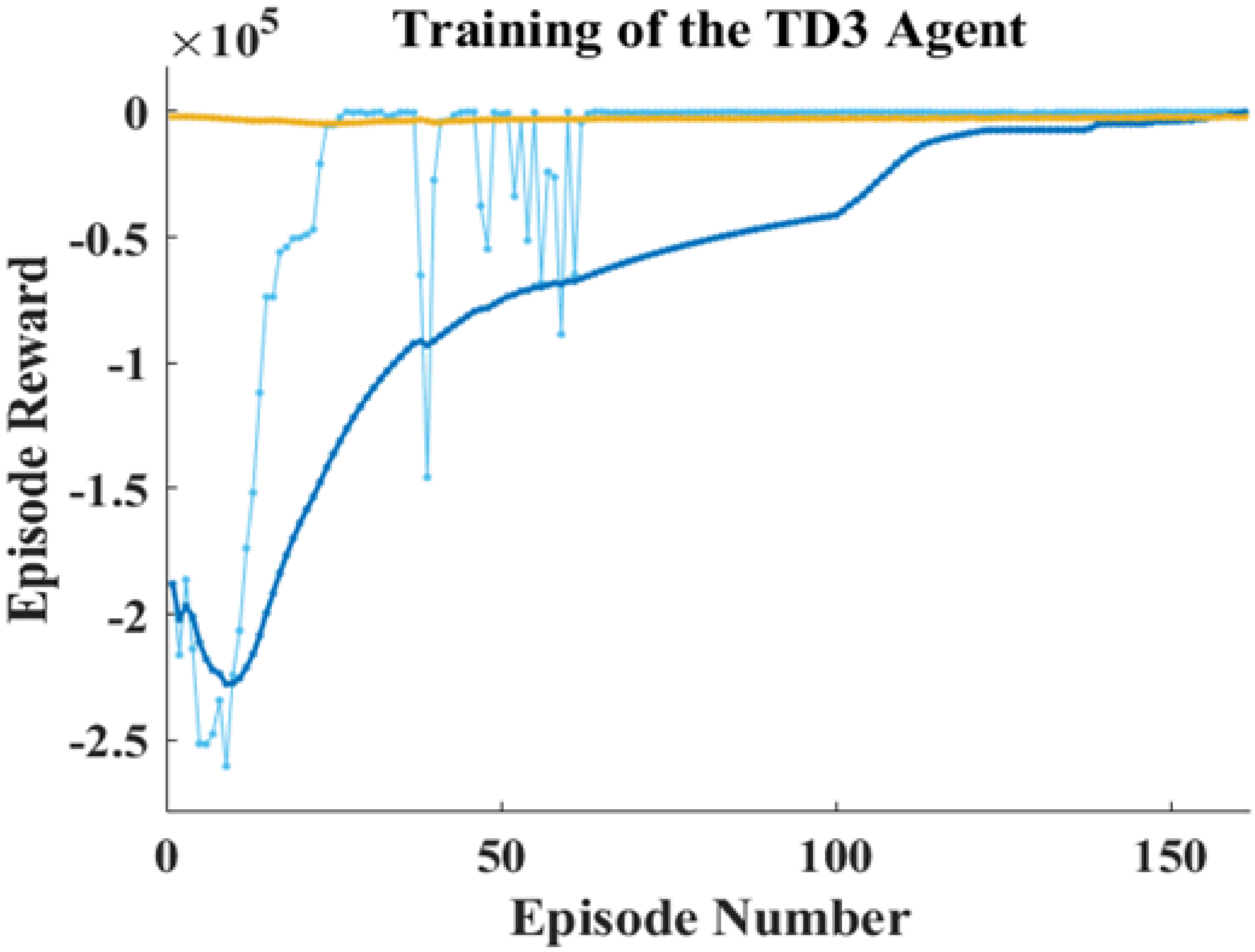
TABLE IV: Training Parameters

Hyperparameter	**Value**
Total Number of Episodes	1000
Maximum steps per episode	100
Actor Learning Rate	5×10^{-4}
Critic Learning Rate	1×10^{-3}
Gradient Threshold	1
Sample Time	0.1 s
Mini-Batch Size	256
Experience Buffer Length	2×10^{6}
Exploration Noise Mean	0
Exploration Noise Standard Deviation	0.15
Standard Deviation Decay Rate	3×10^{-3}

Block Diagram

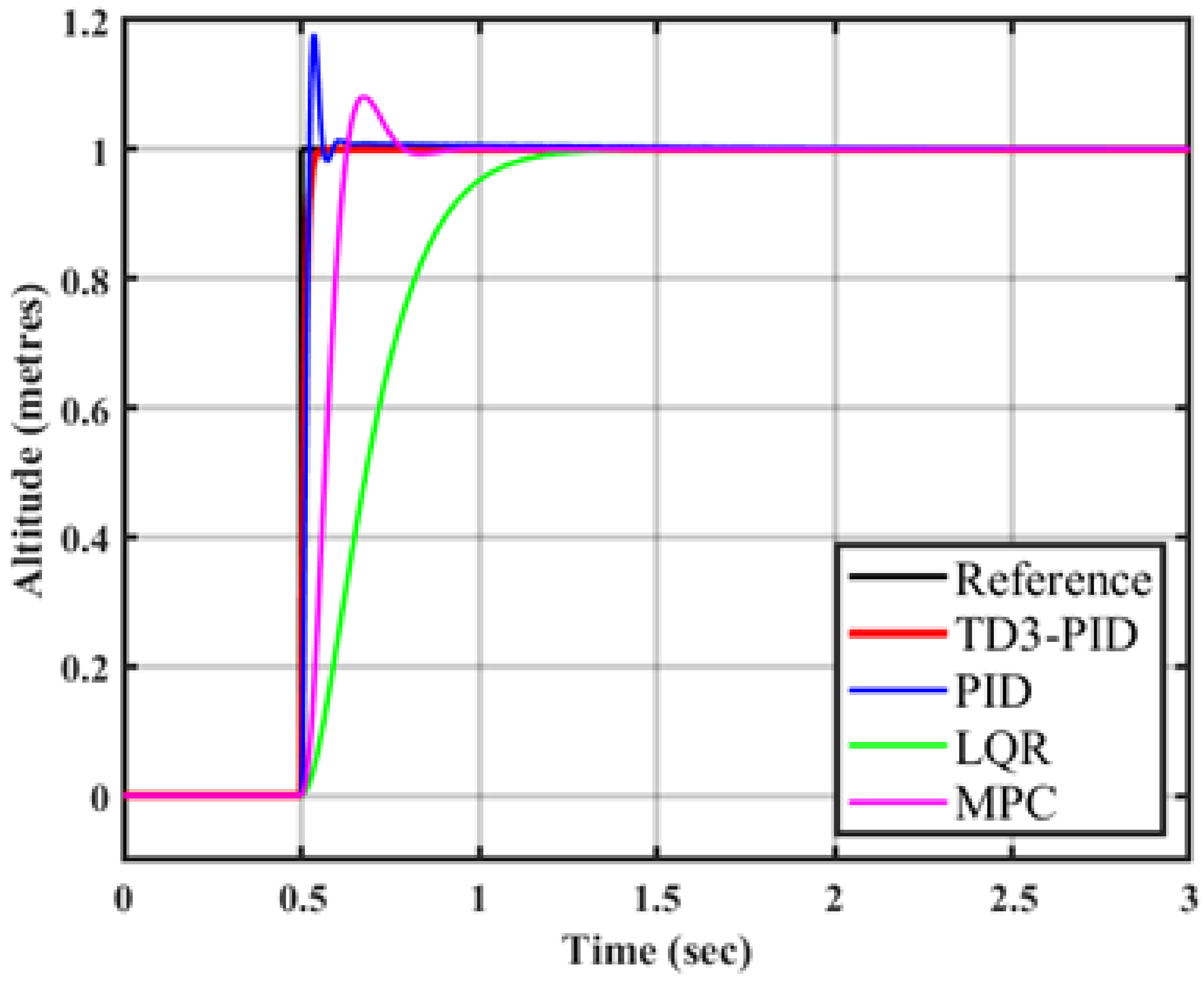


Training

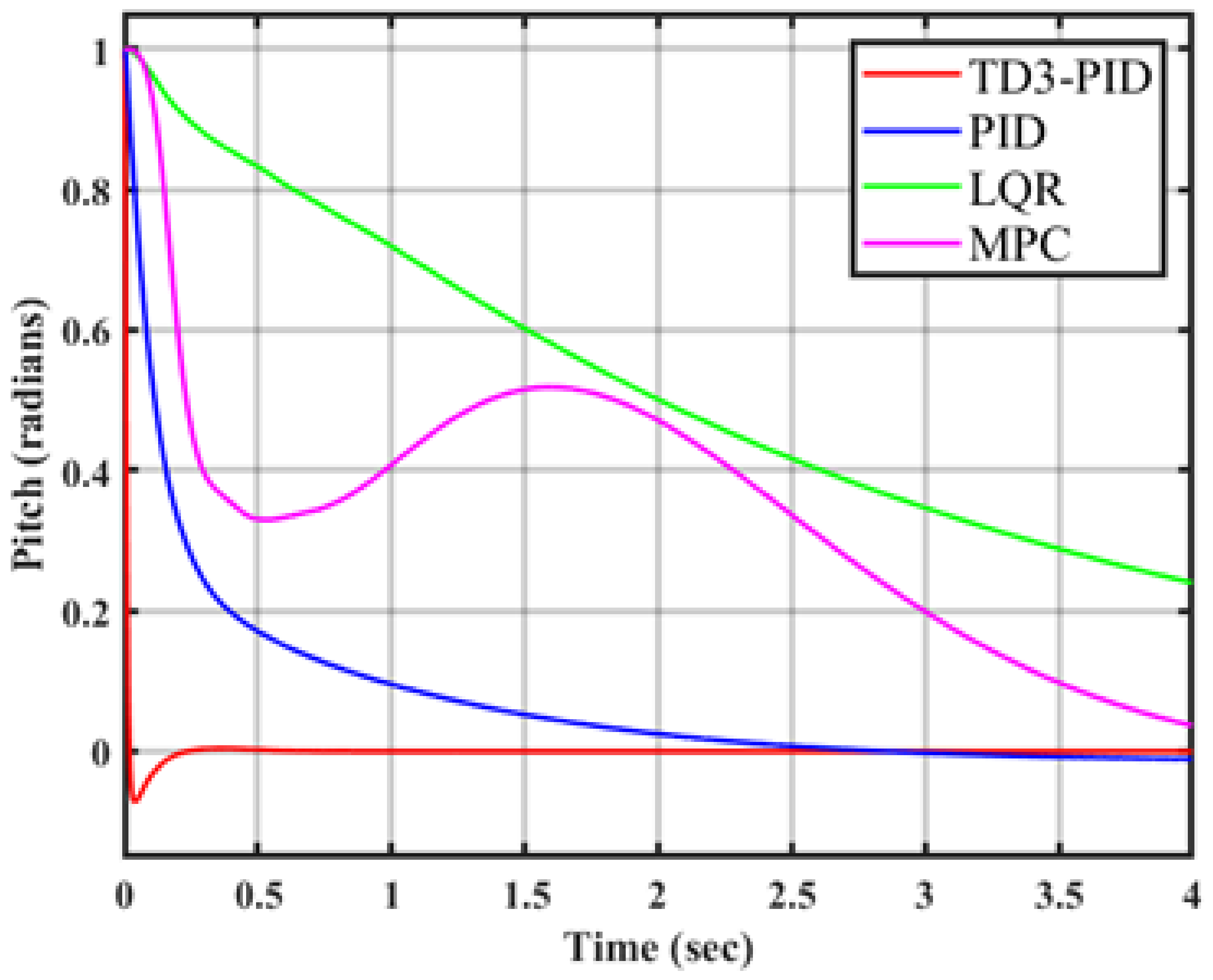


Training of the TD3 Agent
×10^5
0
-0.5
-1
-1.5
-2
-2.5
Episode Reward
0
50
100
150
Episode Number

Tracking Performance

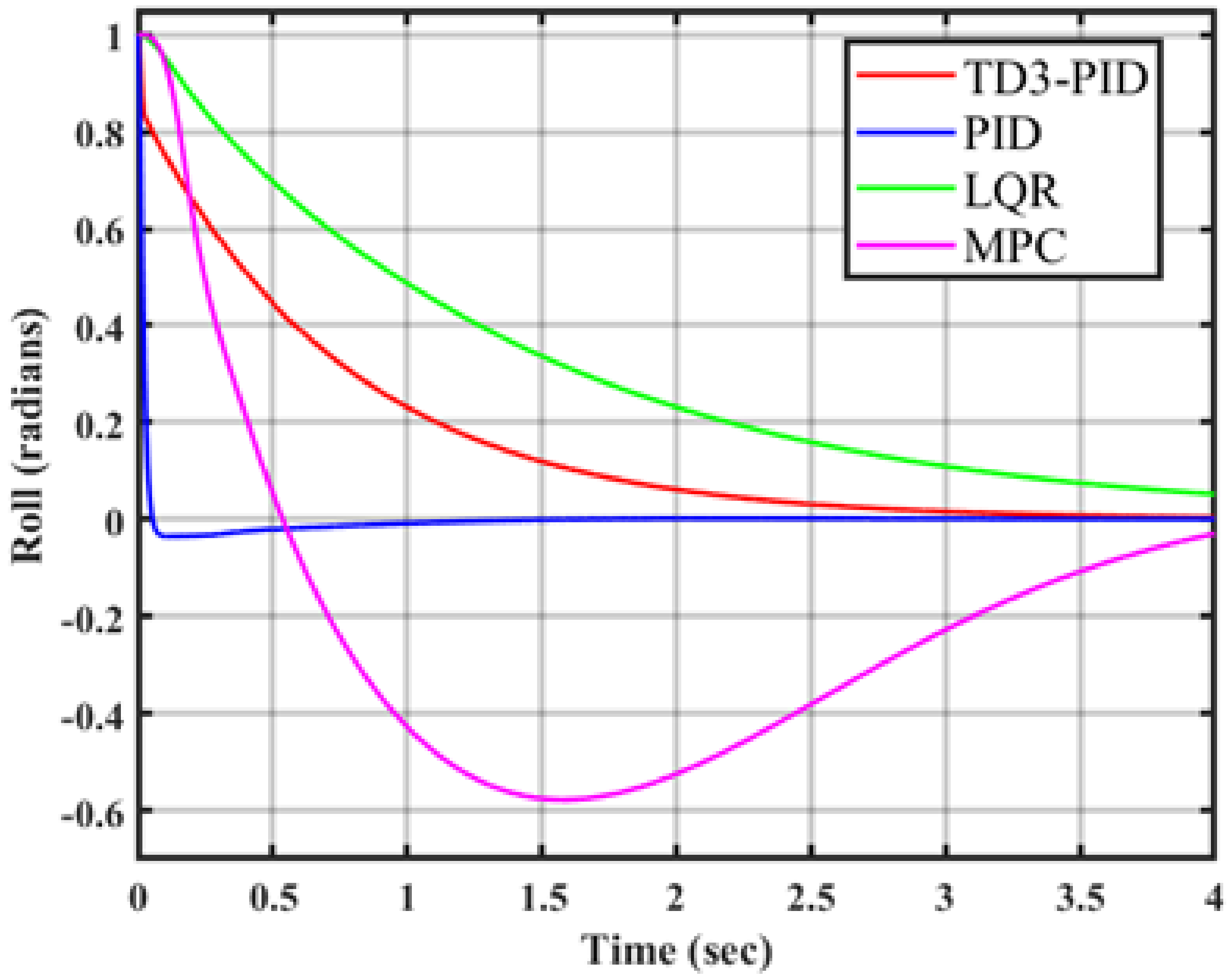


Altitude (m)

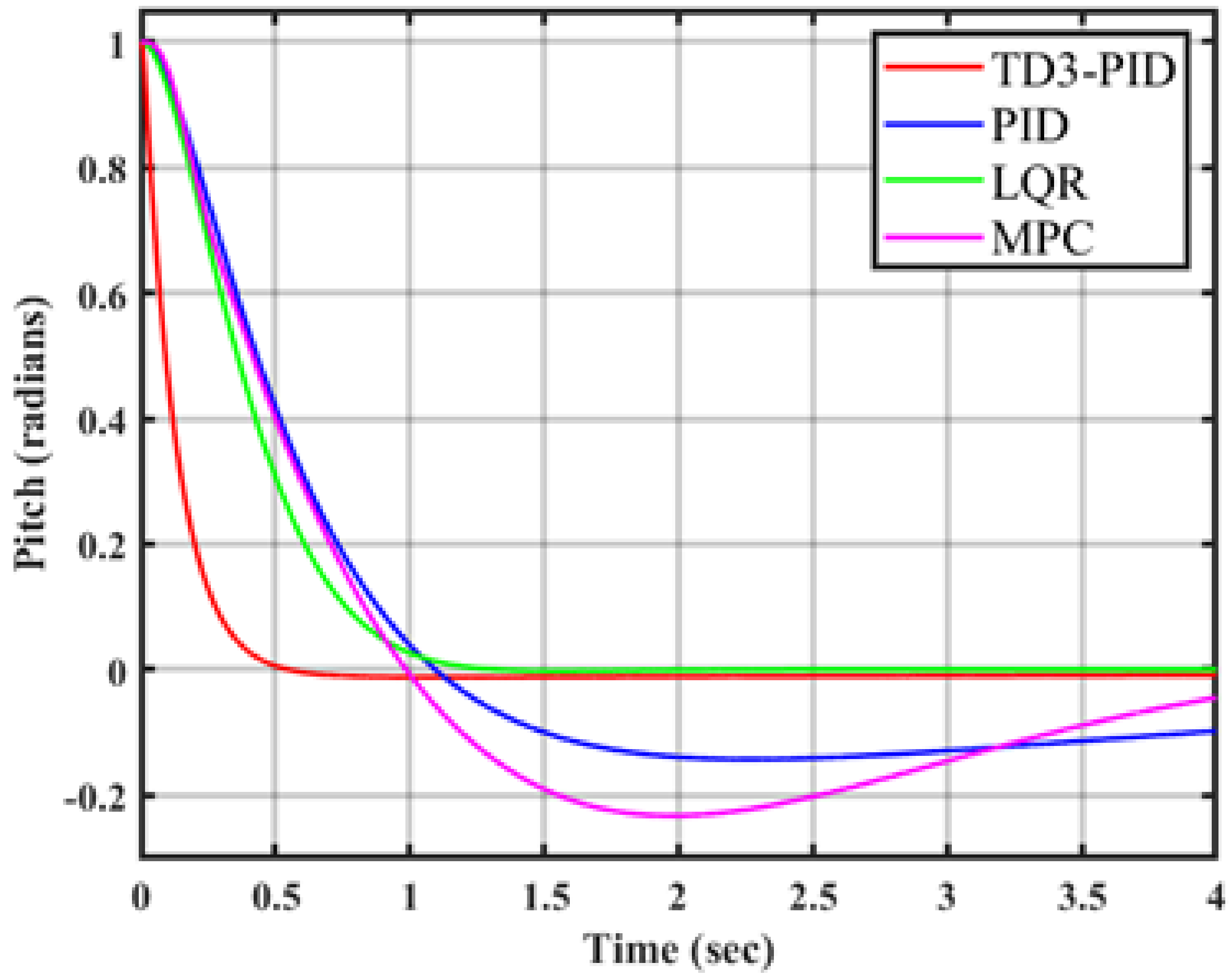


Pitch (radians)

Tracking Performance

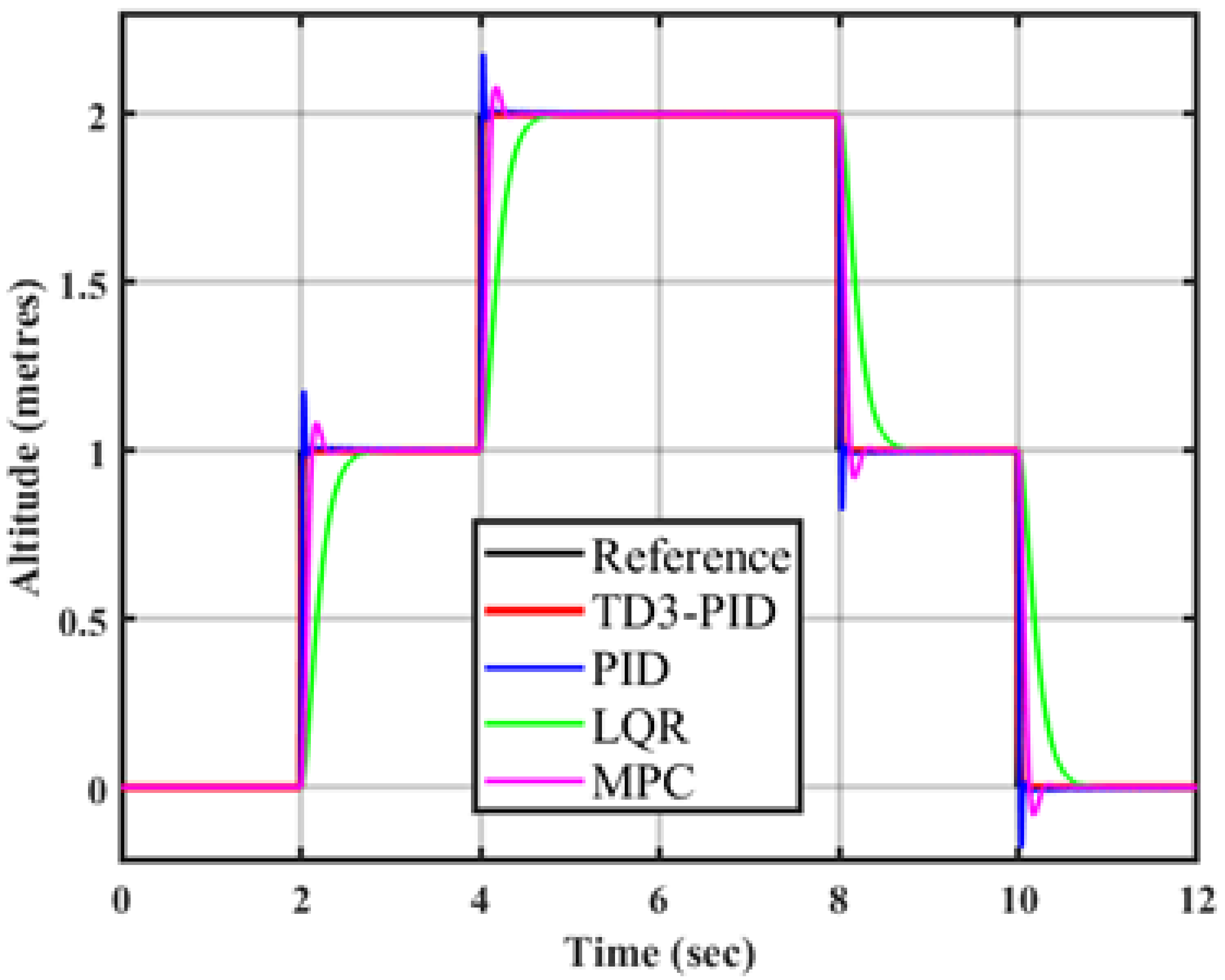


Roll (radians)

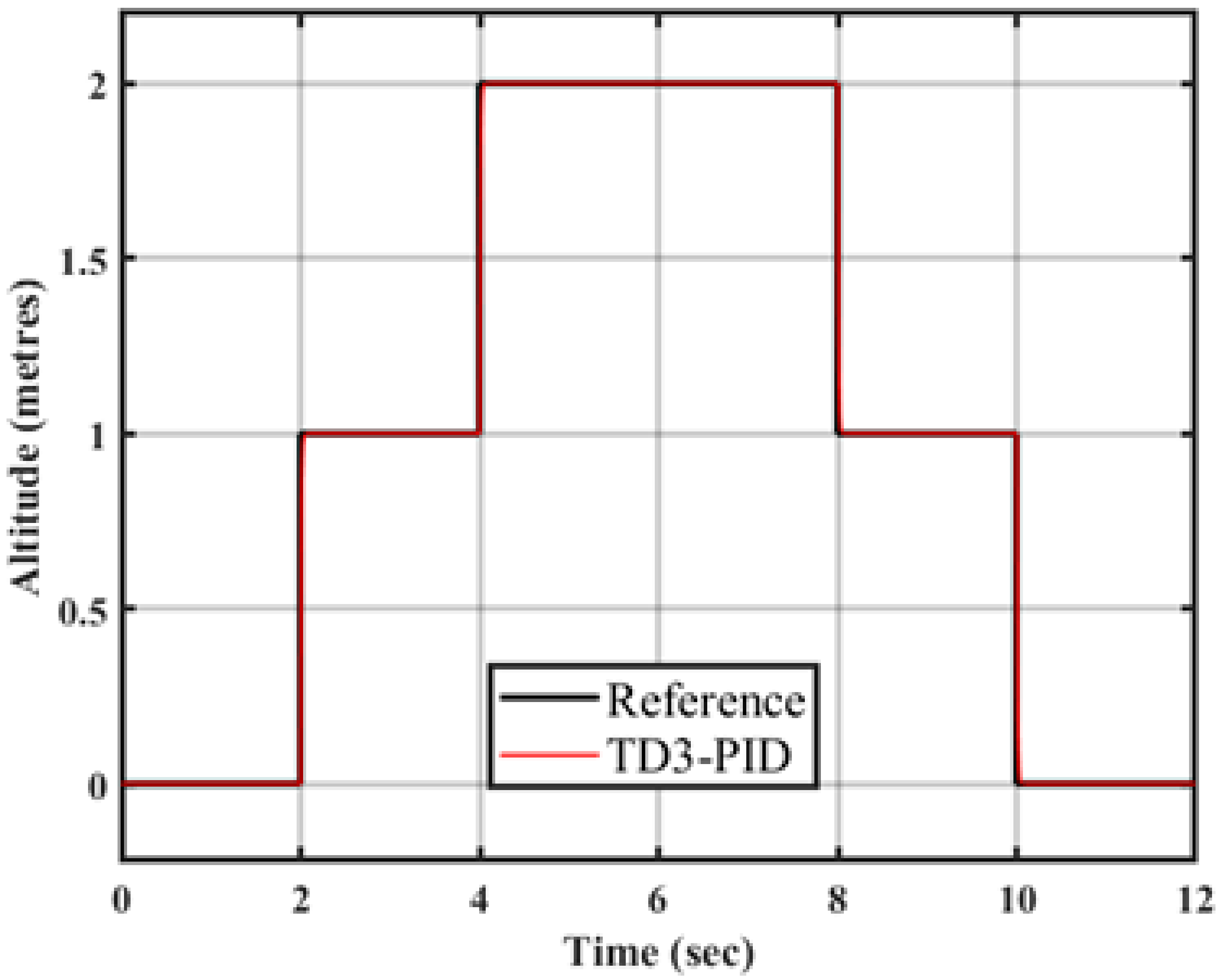


Yaw (radians)

Tracking Performance

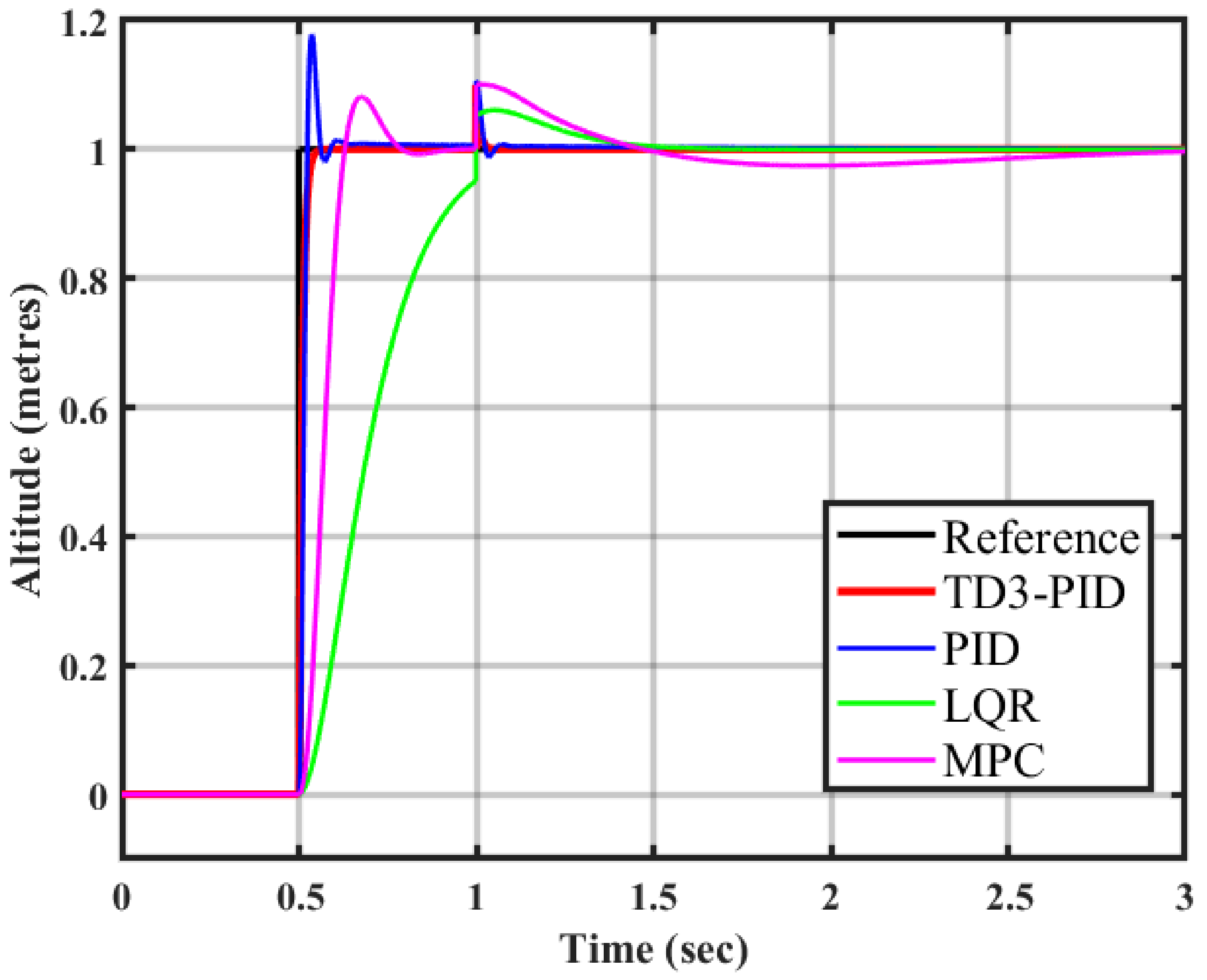


Square wave tracking

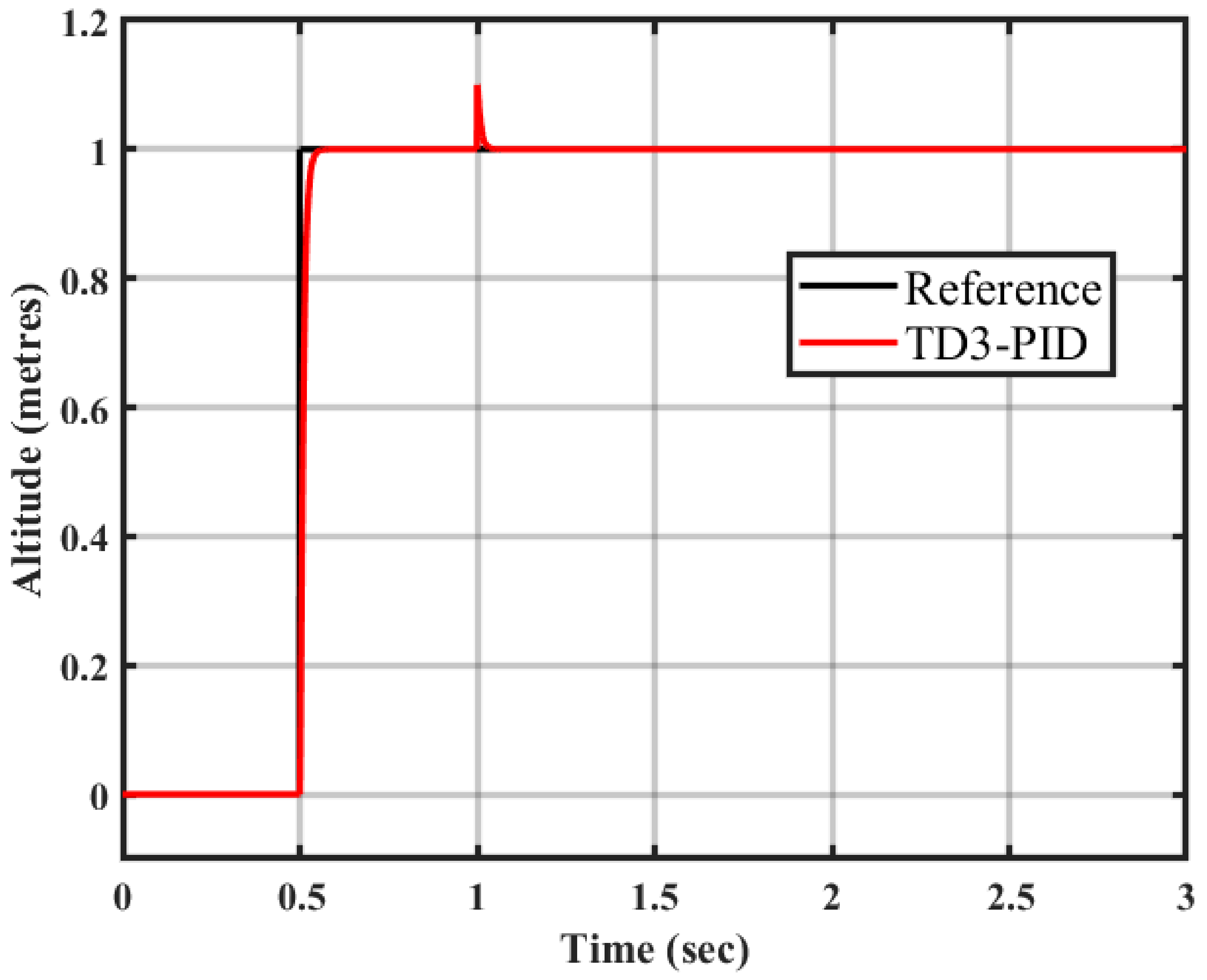


Square wave tracking of the proposed controller

Disturbance Rejection



Disturbance Rejection



Disturbance Rejection of the proposed controller

Time Domain Metrics

Time Domain Parameters	PID	LQR	MPC	TD3-PID
Settling time (sec)	1.0	0.6008	0.2548	0.0354
Rise Time (sec)	0.0161	0.3527	0.0828	0.0195
Overshoot (%)	17.84	0.0013	8.1578	0
Peak Time (sec)	0.036	1.297	0.1766	12
Peak	1.1783	1.00	1.0816	1.00
Settling min	0.9027	0.9005	0.9031	0.9057
Settling max	1.1783	1.00	1.0186	1.00

Performance Metrics

Performance Metric	PID	LQR	MPC	TD3-PID
RMSE	$2.86 * 10^{-6}$	$1.25 * 10^{-17}$	$4.16 * 10^{-17}$	$1.97*10^{-10}$
IAE	0.08148	0.2138	0.075	0.03
ISE	0.0905	0.1357	0.051	0.05067

Results and Discussions

- The TD3-tuned adaptive PID controller consistently outperforms traditional fixed-gain PID in tracking attitude and altitude commands, especially under nonlinear, time-varying, and uncertain conditions.
- Significant reductions in RMSE, IAE, and ISE across test scenarios demonstrate improved precision and robustness in altitude tracking.
- Real-time gain adjustment by the TD3 agent enables quick adaptation during transient flight phases, where conventional PID struggles due to its static gain structure.
- While feedback linearization simplifies the control problem, it remains vulnerable to modeling inaccuracies; the TD3 agent compensates for these residual uncertainties through adaptive gain tuning.
- Controller performance is sensitive to the training environment's diversity and the structure of the reward function; generalization and stability during extreme flight conditions remain challenging.
- Initial tests on Parrot Mambo Minidrones are planned to validate the approach, with focus on minimizing inference latency, ensuring onboard stability, and handling sensor noise within tight computational limits.

Learning Outcomes

- Through this research project, I dove deeper into the various intricacies of mathematical modelling of systems.
- Understanding the dynamics of the quadrotor was challenging in the first phase of development.
- Deriving inverse dynamical equations for feedback linearization was tricky but very rewarding. Understood the practical aspects of feedback linearization.
- Understood the implementation of reinforcement learning in the field of adaptive and learning-based control.
- I successfully bridged concepts from nonlinear control and deep reinforcement learning to address real-world control challenges in aerial robotics.
- I identified limitations of fixed-gain approaches and successfully devised a data-driven adaptive alternative suitable for complex environments.
- I built and validated the complete control pipeline in MATLAB/Simulink, simulating dynamic flight scenarios and disturbance conditions to thoroughly evaluate the performance of the proposed controller.

References

- T. Shuprajhaa, Shiva Kanth Sujit, K. Srinivasan “Reinforcement learning based adaptive PID controller design for control of linear/nonlinear unstable processes ” Applied Soft Computing,Volume 128, 2022, 109450, ISSN 1568-4946
- Bartomeu Rub´i and Bernardo Morcego and Ramon Perez, "A Deep Reinforcement Learning Approach for Path Following on a Quadrotor" in 2020 European Control Conference (ECC), Saint
- Subbareddy Chitta and Ramakalyan Ayyagari, "On the Nonlinear Control of a Class of Cruise Missiles," International Journal of Control, Automation, and Systems, vol. 23, no. 3, pp. 788-797, 2025.
- U. Alejandro-Sanjines, A. Maisincho-Jivaja, V. Asanza, L. L. Lorente-Leyva, and D. H. Peluffo-Ordóñez, “Adaptive PI Controller Based on a Reinforcement Learning Algorithm for Speed Control of a DC Motor,” Biomimetics, vol. 8, no. 5, p. 434, Sep. 2023
- Fujimoto, S., van Hoof, H., \& Meger, D. (2018). “Addressing Function Approximation Error in Actor-Critic Methods.” Proceedings of the 35th International Conference on Machine Learning, 2018.
- Simulink diagrams and MATLAB codes for the TD3 agent are available on the following GitHub link: https://github.com/shreehank22/Adaptive-PID-Control-for-Quadrotor-UAVs-Using-Feedback-Linearization-and-Deep-Reinforcement-Learning