Proceedings of the Seventh International Conference on Research in Intelligent and Computing in Engineering pp. 29–32
DOI: 10.15439/2022R52
ISSN 2300-5963 ACSIS, Vol. 33

Optimal tracking controllers with Off-policy Reinforcement Learning Algorithm in Quadrotor

Dinh Duong Pham, Thanh Trung Cao, Tat Chung Nguyen, Phuong Nam Dao

***Abstract*—In this study, the optimal tracking control problem for the quadrotor which is a highly coupling system with completely unknown dynamics is addressed based on data by introducing the reinforcement learning (RL) technique. The proposed Off-policy RL algorithm does not need any knowledge of quadrotor model. By collecting data, which is the states of quadrotor system then using an actor-critic networks (NNs) to solve the optimal tracking trajectory problem. Finally, simulation results are provided to illustrate the effectiveness of proposed method.**

I. Introduction

In recent years, unmanned aerial vehicle (UAV) has been gaining an increasing consideration in research society due to its huge potential in many areas where the appearance of human is hard to achieve, such as: disaster surveillance, agricultural applications,... One of the most effective UAV is quadrotor with the ability of vertically taking off and landing, the versatile adaptation to arbitrary trajectories. So that, the needs of solving the problem of tracking control for a quadrotor has been put under research for years. Moreover, in application, it could be impossible to have the fully knowledge of the system due to the uncertainty of the environment where the quadrotor functions, the unknown loads that quadrotor carries. So the uncertainty is an indispensable part when it comes to control a quadrotor. Recently, many proposed controllers were proposed: PID-controller [5],[6], Linear-quadratic regulator (LQR) Controller [3],[7], Backstepping technique [1],[2], Sliding mode controller [4]. However, they are not considered as optimal control. In [9], this paper proposed a formation optimal control for multiple quadrotors. But the disadvantages of this scheme is that it just tackled the tracking problem for a simple trajectory which is straight line. This paper proposed a control scheme based on Off-policy Reinforcement Learning algorithm to obtain the optimal controllers for tracking problem of a quadrotor with completely unknown knowledge of the system.

II. Preliminaries and Problem Statement

In this section, we present the model of quadrotor and the traditional control scheme. A quadrotor 1 could be described with dynamic equations:

$$\begin{aligned} m\ddot{p} &= T_p R e_{3,3} - mge_{3,3} \\ J\ddot{\Theta} &= \tau - C(\Theta,\dot{\Theta})\dot{\Theta} \end{aligned} \tag{1}$$

Corresponding Author: Thanh Trung Cao, trung.caothanh@hust.edu.vn, nam.daophuong@hust.edu.vn

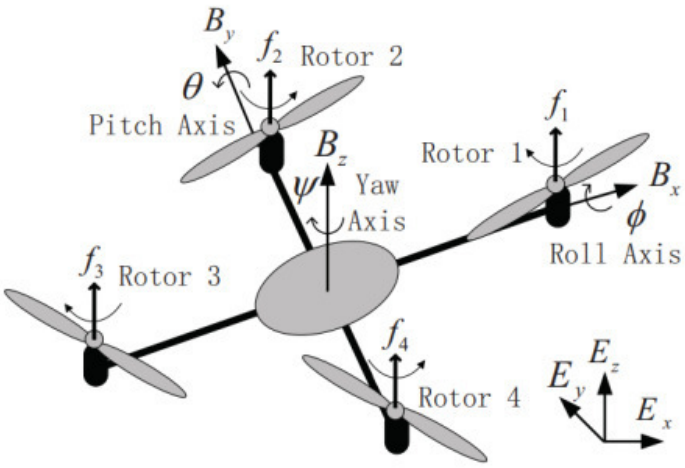

Figure 1. A typical quadrotor

Where:
The position of the center of mass is $p = [p_x, p_y, p_z]^T \in \mathbb{R}^3$. The Euler angles $\Theta = [\phi, \theta, \psi]$. $e_{i,j}$ is the vector which has i numbers of zeros except for number 1 in the j^{th} position.

$C(\Theta,\dot{\Theta}) = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$

We define $T_p \in \mathbb{R}$ is the total force of all propellers $T_p = T_1 + T_2 + T_3 + T_4$ và $\tau = [\tau_\phi, \tau_\theta, \tau_\psi]^T \in \mathbb{R}^3$ is momentum that acts on quadrotor which resolve around x, y, z axis, $T_p = k_w u_z$ and $\tau = [l_\tau k_w u_\phi,\ l_\tau k_w u_\theta,\ k_t u_\psi]^T$. m is the weight, g is gravity, $J = diag(J_x, J_y, J_z)$ with $J_x,\ J_y,\ J_z$ are moments of inertia which resolve around $x,\ y,\ z$ axis respectively. It can be seen that this is 6 DOF which is highly coupling with 4 inputs $u_z, u_\phi, u_\theta, u_\psi$. In fact, $u_z, u_\phi, u_\theta, u_\psi$ are dependent on propellers' velocity:

$$\begin{aligned} u_z &= \omega_1^2 + \omega_2^2 + \omega_3^2 + \omega_4^2 \\ u_\phi &= \omega_2^2 - \omega_4^2 \\ u_\theta &= \omega_1^2 - \omega_3^2 \\ u_\psi &= \omega_1^2 - \omega_2^2 + \omega_3^2 - \omega_4^2 \end{aligned} \tag{2}$$

Figure 2 illustrates the typical cascade control strategy for a quadrotor which consists of position controller in the outer loop and attitude controller in the inner loop.

©PTI 2022

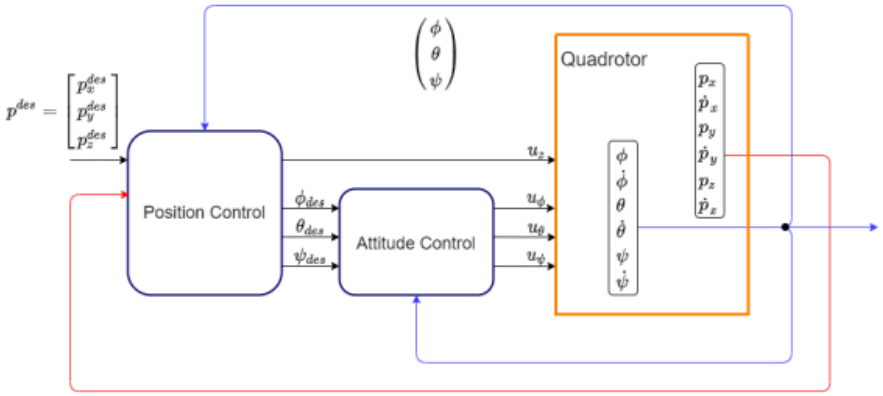

Figure 2. The principle of controlling quadrotors

Remark 1: The control objective is to obtain the optimal tracking control by the Off-policy Algorithm RL for an unknown dynamics quadrotor. It is worth noting that this strategy has the advantage of solving the optimal tracking control problem for a sophisticated trajectory with no prior knowledge about the system by the iterative algorithm to estimate the optimal controller.

III. Proposed Control Strategy

A. Position Controller with Off-policy RL

The position dynamic can be written:

$$\begin{aligned}\ddot{p} &= m^{-1}k_w u_z R e_{3,3} - g e_{3,3} \\ &= m^{-1}k_w u_p\end{aligned} \tag{3}$$

Note that $u_p = u_z R e_{3,3} - \frac{m}{k_w} g e_{3,3} \in \mathbb{R}^3$ Set $x_p = [p_x, \dot{p}_x, p_y, \dot{p}_y, p_z, \dot{p}_z]^T \in \mathbb{R}$. $A_p = diag(a_p, a_p, a_p) \in \mathbb{R}^{6\times 6}$, $a_p = [0_{2\times 1}\ e_{2,1}]$ and $B_p = m^{-1}k_w[e_{6,2}, e_{6,4}, e_{6,6}]$ Assume the desired trajectory $\dot{x}_{pd}$ has $\dot{x}_{pd} = A_{pd}x_{pd}$ and error $e_p = x_p - x_{pd}$, we could rewrite the expanded system:

$$\dot{X}_p = \begin{bmatrix}\dot{e}_p \\ \dot{x}_{pd}\end{bmatrix} = \begin{bmatrix}A_p & A_p - A_{pd} \\ 0_{6\times 6} & A_{pd}\end{bmatrix} X_p + \begin{bmatrix}B_p \\ 0_{6\times 3}\end{bmatrix} u_p \tag{4}$$

The cost function is chosen:

$$V_p(X_p(t)) = \int_t^\infty e^{-\lambda(\tau - t)}(X_p(\tau)^T Q_p X_p(\tau) + u_p(\tau)^T R_p u_p(\tau))d\tau \tag{5}$$

The Off-policy RL algorithm [8] for this optimal control problem is proposed as:

1) Initiate:
   Start with an acceptable control input u_p^0 and noise u_{pe} which is added to guarantee PE condition. Collect data and determine a threshold ϵ_p
2) Policy Evaluation
   With $u_p^i(X_p)$ solved from previous iteration, let solve $V_p^{i+1}(X_p)$ và $u_p^{i+1}(X_p)$ from equation:

$$\begin{aligned}&V_p^{i+1}(X_p(t+\delta t)) - V_p^{i+1}(X_p(t)) \\ &= -\int_t^{t+\delta t}[X_p(\tau)^T Q_p X_p(\tau) \\ &\quad + [u_p^i(X_p(\tau))]^T R_p u_p^i(X_p(\tau))]d\tau \\ &\quad + \int_t^{t+\delta t}\lambda V_p^{i+1}(X_p(\tau))d\tau \\ &\quad + 2\int_t^{t+\delta t}[u_p^{i+1}(X_p(\tau))]^T R_p u_p^i(X_p(\tau))d\tau \\ &\quad - 2\int_t^{t+\delta t}[u_p^{i+1}(X_p(\tau))]^T R_p[u_p^0(\tau) + u_{pe}]d\tau\end{aligned} \tag{6}$$

3) Policy Improvement
   Continue to iterate until $\|u_p^{i+1} - u_p^i\| < \epsilon_p$

To approximate V_p^i và u_p^i, Critic-Actor NNs ware estimated as:

$$V_p^i(X_p) = w_{Vp}^T \varphi_p(X_p) \tag{7}$$

$$u_p^i(X_p) = w_{up}^T \psi_p(X_p) \tag{8}$$

In detail, $\varphi_p(X_p) \in \mathbb{R}^{l_1}$ and $\psi_p(X_p) \in \mathbb{R}^{l_2}$ are 2 activation function vectors. $w_{Vp} \in \mathbb{R}^{l_1\times 1}$ and $w_{up} \in \mathbb{R}^{l_2\times 3}$ are 2 weight vectors respectively. Then, we can apply Least-Square Algorithm to solve (6).
After obtaining optimal $u_p = [u_{px}, u_{py}, u_{pz}]^T$, we could have u_z, desired attitudes angles:

$$\begin{aligned}u_z &= \sqrt{u_{px}^2 + u_{py}^2 + (u_{pz} + u_b)^2} \\ \psi_d &= 0 \\ \phi_d &= arcsin(\frac{u_{px}sin(\psi_d) - u_{py}cos(\psi_d)}{u_z}) \\ \theta_d &= arctan(\frac{u_{px}cos(\psi_d) + u_{py}sin(\psi_d)}{u_{pz} + u_b})\end{aligned} \tag{9}$$

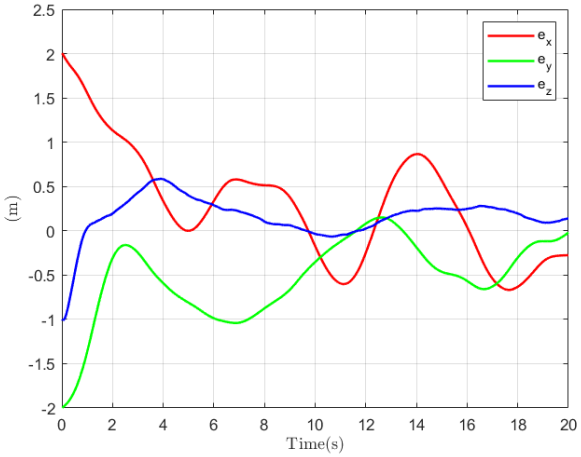

Figure 3. The position tracking error at initial stage

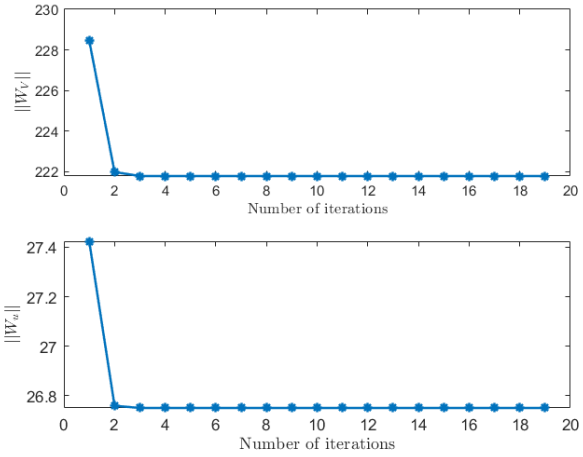

Figure 4. The convergence of weights in position controller

B. *Attitude Controller with Off-policy RL*

In the inner loop, similar control structure is proposed. Set $x_\Theta = [\phi, \dot{\phi}, \theta, \dot{\theta}, \psi, \dot{\psi}]^T$.Here, we have $\dot{x}_\Theta = F_\Theta x_\Theta + B_\Theta u_\Theta$ and $B_\Theta = [e_{6,2}b_{\Theta 1}, e_{6,4}b_{\Theta 2}, e_{6,6}b_{\Theta 3}] \in \mathbb{R}^{6\times 3}$ $b_{\Theta 1} = J_x^{-1} l_\tau k_w$, $b_{\Theta 2} = J_y^{-1} l_\tau k_w$, $b_{\Theta 3} = J_z^{-1} k_t$. The desired trajectory of attitude was obtained in the outer loop, which could be described as $\dot{x}_{\Theta d} = F_{\Theta d} x_{\Theta d}$. Let have $e_\Theta = x_\Theta - x_{\Theta d}$. Firstly, we have the expanded system:

$$\dot{X}_{\Theta d} = \begin{bmatrix} \dot{e}_\Theta \\ \dot{x}_{\Theta d} \end{bmatrix} = \begin{bmatrix} F_\Theta & F_\Theta - F_{\Theta d} \\ 0_{6\times 6} & F_{\Theta d} \end{bmatrix} X_{\Theta d} + \begin{bmatrix} B_\Theta \\ 0_{6\times 3} \end{bmatrix} u_\Theta \tag{10}$$

The cost function is chosen as:

$$V_\Theta(X_\Theta(t)) = \int_t^\infty e^{-\lambda(\tau - t)} (X_\Theta(\tau)^T Q_e X_\Theta(\tau) + u_\Theta(\tau)^T R u_\Theta(\tau)) d\tau \tag{11}$$

Then, we continue to implement the iterative algorithm as in the previous section.

IV. Simulation

Consider a quadrotor with the desired trajectory is a spiral trajectory

At the first stage, we use 2 PID-controllers for both outer and inner loops to collect data for the next stage of training to obtain the optimal controllers. Note that noises is added to the system to guarantee the PE condition.

The position tracking error in this stage is illustrate in Fig 3.

Then, we use the data as the input to the algorithms which are proposed in the previous section. The convergences of the weights are shown in Fig 4.

After we obtain the weights, estimated optimal controllers are applied to the object. The tracking performance is illustrated in Fig 5.

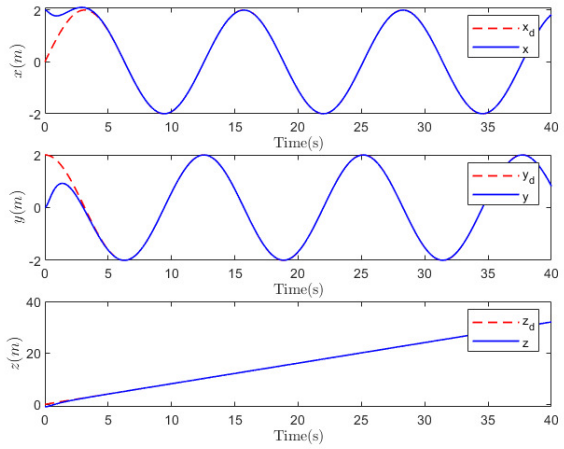

Figure 5. The tracking position of the optimal controllers

V. Conclusion

In this paper, a novel control strategy which consists of the Off-policy RL algorithm was proposed. By collecting data to train two actor-critic networks (NNs) which aim to estimate the optimal controllers which includes position controller and attitude controller, this structure has the advantage of no need of any prior information of the highly coupling system. Finally, simulation results are provided to illustrate the tracking performance of a sophisticated trajectory of the system.

Acknowledgement(s)

Dinh Duong Pham was funded by Vingroup JSC and supported by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Institute of Big Data, code VINIF.2021.ThS.43.

REFERENCES

[1] E. Altug, J. Ostrowski, and R. Mahony. Control of a quadrotor helicopter using visual feedback. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, volume 1, pages 72–77 vol.1, 2002.

[2] A. Das, F. Lewis, and K. Subbarao. Backstepping approach for controlling a quadrotor using lagrange form dynamics. *Journal of Intelligent and Robotic Systems*, 56:127–151, 09 2009.

[3] Y. Li and S. Song. A survey of control algorithms for quadrotor unmanned helicopter. In *2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*, pages 365–369, 2012.

[4] C. Mu, C. Sun, and W. Xu. Fast sliding mode control on air-breathing hypersonic vehicles with transient response analysis. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 230, 11 2015.

[5] A. Tayebi and S. McGilvray. Attitude stabilization of a vtol quadrotor aircraft. *IEEE Transactions on Control Systems Technology*, 14(3):562–571, 2006.

[6] A. Tayebi and S. McGilvray. Attitude stabilization of a vtol quadrotor aircraft. *IEEE Transactions on Control Systems Technology*, 14(3):562–571, 2006.

[7] W. Wang, H. Ma, and C.-Y. Sun. Control system design for multi-rotor mav. *Journal of Theoretical and Applied Mechanics*, 51:1027–1038, 01 2013.

[8] G. Xiao, H. Zhang, Y. Luo, and H. Jiang. Data-driven optimal tracking control for a class of affine non-linear continuous-time systems with completely unknown dynamics. *Iet Control Theory and Applications*, 10:700–710, 2016.

[9] W. Zhao, H. Liu, F. L. Lewis, and X. Wang. Data-driven optimal formation control for quadrotor team with unknown dynamics. *IEEE Transactions on Cybernetics*, pages 1–10, 2021.