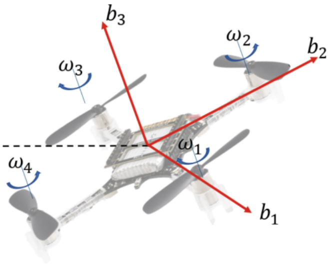
Safe reinforcement learning-based tracking control with application to quadrotor obstacle avoidance

Junkai Tan

Electrical Engineering, Xi'an Jiaotong University

2024.4.9

Preliminaries: quadrotor dynamics



According to Newton's Second Law, the equilibrium of forces is modeled as:

$$ {}^b\dot{v} = -{}^b\omega \times {}^b v + \frac{{}^bF}{m} + \frac{R_n^b e_3 g}{m} \qquad (1) $$

Assuming that the angle is small enough, dynamics can be abbreviated as:

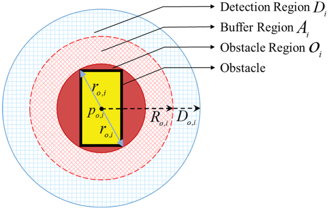
$$ \begin{aligned} {}^n\dot{v}_h &= \frac{f}{m}\begin{bmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{bmatrix}\begin{bmatrix} \theta \\ \varphi \end{bmatrix} = \frac{f}{m}A_\psi \Theta_h \\ {}^n\dot{v}_z &= g + \frac{f}{m} \end{aligned} \qquad (2) $$

where ${}^*p$ is the position of the quadrotor, ${}^*\Omega$ is the position, ${}^*v$ is the velocity, and ${}^*\omega$ is the angular velocity.

Preliminaries: obstacle modeling

For the obstacle modeling, three regions of the obstacle $x_{o,i}$ are defined:

1. Detection Region $\mathcal{D}_i$: Triggering the execution of obstacle avoidance strategy.
2. Buffer Region $\mathcal{A}_i$: Buffer layer for possible upcoming collisions with obstacles.
3. Obstacle Region $\mathcal{O}_i$:Colliding with the obstacle after entering this region.



Design function $s_i(x)$ to characterize the regions of obstacle $x_{o,i}$:

$$s_i(x) = \begin{cases} 0, & d_{o,i} > D_{o,i}, \\ l_1 + l_1 \cos(\pi \frac{d_{o,i}^2 - R_{o,i}^2}{D_{o,i}^2 - R_i^2}), & R_{o,i} < d_{o,i} \leqslant D_{o,i}, \\ l_2 + l_3 \cos(\pi \frac{d_{o,i}^2 - r_{o,i}^2}{R_{o,i}^2 - r_{o,i}^2}), & r_{o,i} < d_{o,i} \leqslant R_{o,i}, \\ 1, & d_{o,i} \leqslant r_{o,i}, \end{cases} \quad \begin{cases} l_2 + l_3 = 1, \\ l_2 - l_3 = 2 l_1 \end{cases}$$

Problem formulation: Control system and objective

Nonlinear tracking control system

Defining the tracking desired trajectory error $e = x_d - x$, then the path tracking model of the quadrotor can be expressed as:

$$\dot{e} = f(e) + \sum_{i=1}^{N} g_i(e) u_i \tag{3}$$

where f and g is the quadrotor dynamics, u_i is the control input.

Performance index

To obtain the tracking controller, design J_i in quadratic form:

$$J_i\left(e_0, u(\cdot)\right) \triangleq \int_t^{\infty} r_i(e(t), u) dt = \int_0^{\infty} (Q_i(e) + \sum_{j=1}^{N} u_j^T R_{ij} u_j) \mathrm{d}t$$

Problem formulation: Finding Nonzero-sum Game Optimal Controller

The following optimization problem is established and theoretically analyzed and solved:

1. Optimization objective: Value Function（价值函数）:

$$V_i^*(x_0) \triangleq \min_{u(\cdot)\in\mathcal{S}(x_0)} J_i(x_0, u(\cdot)), \quad x_0 \in \mathbb{R}^n \tag{4}$$

2. Condition for optimization : Hamiltanian（汉密尔顿算子）:

$$H\left(e, u, V_i'^{\mathrm{T}}\right) \triangleq r_i(e,u) + (\Delta V_i^{*\ T}(f + \textstyle\sum_{j=1}^N g_j u_i) \tag{5}$$

3. Pontryagin's maximum theory-based optimization solution:

$$u_i^\star(e) \triangleq \arg\min_{u\in\mathbb{R}^m} H\left(e, u_i, V_i'^{\mathrm{T}}\right) = -\frac{1}{2} R_{ii}^{-1} g_i^T (\Delta V_i)^T \tag{6}$$

Safe RL tracking control: Definitions

In the quadrotor control application, controllers that lack security are generally difficult to apply, so this research will theoretically analyze and design a safe RL-based controller

1. Definition: Safety Region c

Define c be the safety state region and $h(x)^a$ be the boundary function:

$$c = \{x \in \mathbb{R}^n \mid h(x) \geq 0\}$$
$$\partial c = \{x \in \mathbb{R}^n \mid h(x) = 0\}$$
$$\mathrm{Int}(c) = \{x \in \mathbb{R}^n \mid h(x) > 0\}$$

[a] Junkai Tan et al. "Nash Equilibrium Solution Based on Safety-Guarding Reinforcement Learning in Nonzero-Sum Game". In: *2023 International Conference on Advanced Robotics and Mechatronics (ICARM)*.

2. Definition: Barrier Function

Design the barrier function $b(x)^a$:

$$b(x) = [\frac{1}{h(x)} - \frac{1}{h(0)}]^2$$

- $\forall x(t) \in Int(c)$, $|b(x)| < \infty$
- $\lim_{x \to \partial c} b(x) = \infty$
- $b(0) = 0$

[a] Junkai Tan et al. "Safe Human-Machine Cooperative Game with Level-k Rationality Modeled Human Impact". In: *2023 IEEE International Conference on Development*

Safe RL tracking control: NN Design

Neural networks design

According to Weierstrass theorem, a neural network (NN) is designed to approximate the value function and controller.

- Approximate value function:
  $V_i(e,x) = W_i^{\star\mathrm{T}}\phi_i(e) + b(x) + \epsilon_i(e)$
- Approximate controller:
  $u_i^{\star} = -\frac{1}{2}R_{ii}^{-1}g_i(e)^{\mathrm{T}}\left(\phi_i'^{\mathrm{T}}(e)W_i^{\star} + b'^{\mathrm{T}}(x) + \epsilon_i'^{\mathrm{T}}(x)\right)$

Optimization objective: value function

Value function $V_i(x_0)$ is the extremum of performance $J_i(x_0, u(\cdot))$:

$$V_i\left(e_0, x_0\right) \triangleq \min_{u(\cdot)} J_i\left(e_0, x_0, u(\cdot)\right) = \min_{u(\cdot)} \int_0^{\infty}\left(Q_i(e) + \sum_{j=1}^{N} u_j^T R_{ij} u_j + b(x)\right)\mathrm{d}t$$

Safe RL tracking control: Online Learning

To realize the online update of NN weights, Hamiltonian error is set here as the base element of the update target

$$\delta_i = \Omega_i^{\mathrm{T}} \sigma_i + x^{\mathrm{T}} Q_i x + \sum_{j=1}^{N} \frac{1}{4} \omega_j^{\mathrm{T}} \sigma_j' G_{ij} \sigma_j'^{\mathrm{T}} \omega_j + \nabla \epsilon_i^T \Omega_i \tag{7}$$

The optimizing object set as normalized least squares Hamiltonian error:

$$E_i = \frac{1}{2} \left[\frac{\sigma_i^2}{\left(1 + \sigma_i^T \sigma_i\right)^2} + \sum_{k=1}^{M} \frac{(\sigma_i^k)^2}{\left(1 + (\sigma_i^k)^T \sigma_i^k\right)^2} \right] \tag{8}$$

The NN is updated both using current data and historical data:

$$\dot{\hat{\omega}}_i = -\beta_i \frac{\partial E_i}{\partial \omega_i} = -\beta_i \frac{\sigma_i e_i}{\left(1 + \sigma_i^T \sigma_i\right)^2} - \beta_i \sum_{k=1}^{M} \frac{\sigma_i^k e_i^k}{\left(1 + (\sigma_i^k)^T \sigma_i^k\right)^2} \tag{9}$$

Safe RL tracking control: Stability Theory

Theorem1: Asymptotic stability

NN weights are asymptotically stable as following conditions are met:

$$\begin{cases} \overline{g}_i\overline{\phi}_j < 0 \\ \rho < 0 \\ \beta_i\left(\frac{p+1}{2} - 2\lambda_{\min}\left(\Gamma_k\right)\right) < 0 \end{cases} \tag{10}$$

where $\rho = \sum_{i=1}^{N}\left[\beta_i\frac{p+1}{2}\overline{\varepsilon}_i^2 - \left(\overline{\omega}_i\overline{\phi}_i + \overline{\epsilon}_i\right)\sum_{j=1}^{N}\left(\frac{1}{2}G_j\overline{\phi}_i\|\hat{\omega}_j\| - g_i\overline{\epsilon}_i\right)\right]$

Proof: Set the Lyapunov function

$$V_L = \sum_{i=1}^{N}\left(V_i + V_{\omega,i}\right) \tag{11}$$

where $V_{\omega,i} = \frac{1}{2}\tilde{\omega}_i^{\mathrm{T}}\tilde{\omega}_i$

Safe RL tracking control: Stability Theory

According to the given assumptions, the following inequality holds:

$$\dot{V}_i \leq -r_i - \left(\overline{\omega}_i \overline{\phi}_i + \overline{\epsilon}_i\right) \sum_{j=1}^{N} \left(\frac{1}{2} G_j \overline{\phi}_i \|\hat{\omega}_j\| - g_i \overline{\epsilon}_i \right) \tag{12}$$

$$\dot{V}_{\omega,i} \leq \beta_i \left[\frac{p+1}{2} - 2\lambda_{\min}\left(\Gamma_k\right) \right] \|\tilde{\omega}_i\|^2 + \beta_i \frac{p+1}{2} \overline{\varepsilon}^2_{\mathrm{hmax},i} \tag{13}$$

$$\begin{aligned} \dot{V} \leq & -\sum_{i=1}^{N} r_i + \rho \\ & + \sum_{i=1}^{N} \left[\overline{g}_i \overline{\phi}_i + \beta_i \left(\frac{p+1}{2} - 2\lambda_{\min}(\Gamma_k) \right) \right] \|\tilde{\omega}_i\|^2 \end{aligned} \tag{14}$$

So $\dot{V}_L \leq 0$ holds, i.e., the asymptotic stability of the weights is proved

Experiment Setup

The detailed experiment setup is listed as follows:

1. Operation platform: Rflysim, Matlab Simulink.
2. Aircraft model: DJI-F450 quadrotor.
3. Control frequency: 30Hz.

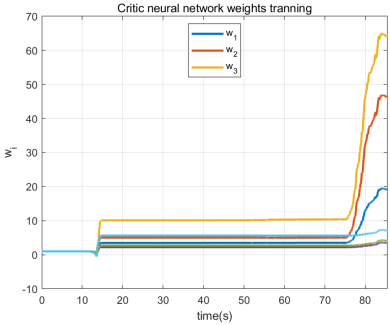


图 1: Learning process of NN

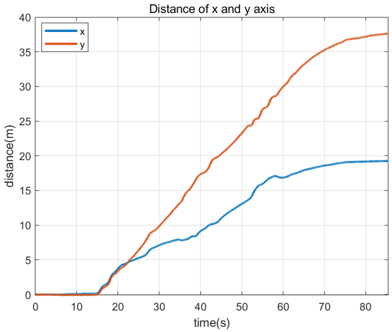


图 2: Position in X-axis and Y-axis

Experiment Results

The trajectories of obstacle avoidance tracking control are shown in fig4.

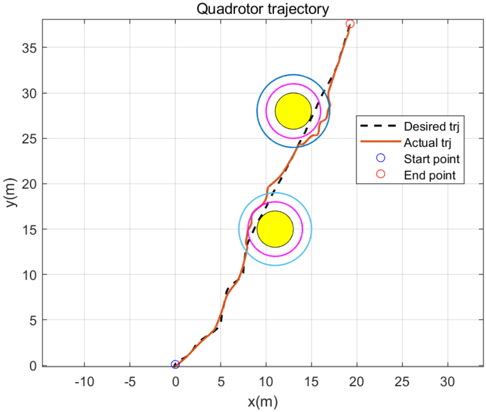


图 3: Quadrotor tracking control with obstacle avoidance

Experiment Setup

The detailed experiment setup is listed as follows:

1. Operation platform: Rflysim motion capture OptiTrack.
2. Aircraft model: Droneyee-X150 quadrotor.
3. Control frequency: 30Hz.

图 4: Droneyee-X150 quadrotor

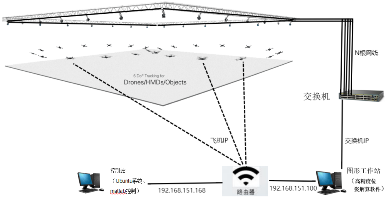


图 5: Motion capture OptiTrack

Experiment Results

The learning process of the NN weight is presented in fig3. The control input to the quadrotor is showed in fig4.

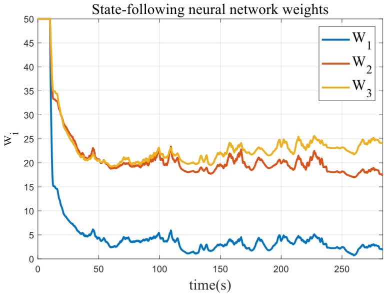


图 6: NN weight learning process

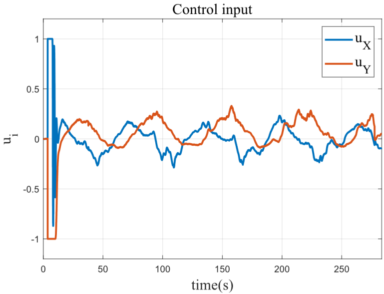


图 7: Control input to the quadrotor

Experiment Results

The positions and tracking error of the quadrotor are presented in fig5 and fig6, respectively.

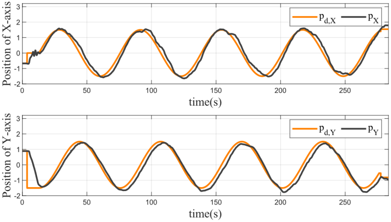


图 8: Desired and actual position of quadrotor

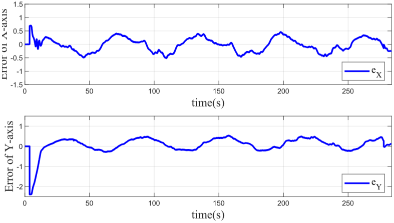


图 9: Tracking error of quadrotor

Experiment Results

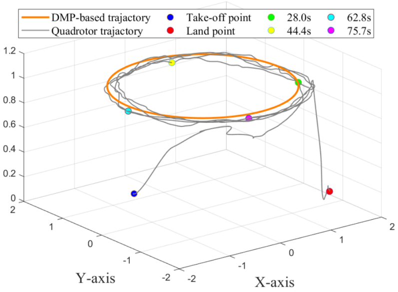


图 10: 3-Dimension trajectories of quadrotor

图 11: Experiment environment

References I

[1] Junkai Tan et al. "Nash Equilibrium Solution Based on Safety-Guarding Reinforcement Learning in Nonzero-Sum Game". In: *2023 International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE. 2023, pp. 630–635.

[2] Junkai Tan et al. "Safe Human-Machine Cooperative Game with Level-k Rationality Modeled Human Impact". In: *2023 IEEE International Conference on Development and Learning (ICDL)*. IEEE. 2023, pp. 188–193.

Thanks!

Thanks for your listening