Adaptive Concurrent Learning Algorithm Based on Pontryagin's Maximum Principle for Nonlinear System Optimal Tracking Control with State Inequality Constraints

Yuqi Zhang
Beijing University of Posts and Telecommunications, School of Artificial Intelligence, Beijing 100876, China

Bin Zhang
Beijing University of Posts and Telecommunications, School of Artificial Intelligence, Beijing 100876, China
Email: zhangbinzdh@bupt.edu.cn

Abstract

In this paper, an adaptive iterative learning algorithm is proposed to solve the optimal tracking control problem (OTCP). Unlike the existing method, we select the finite-time horizon cost function to measure tracking performance. Our method doesn't require an accurate system dynamic and the concurrent learning (CL) technique is utilized to learn the system identification model. Based on identification model, we present our concurrent iterative learning algorithm under the Pontryagin's framework to learn the solution for OTCP. The proposed algorithm overcomes the limitation of Adaptive Dynamic Programming (ADP) methods when dealing with time-varying systems or situations involving state inequality constraints. The algorithm's effectiveness is demonstrated through a numerical simulation.

Keywords: Optimal tracking control, Concurrent learning, State inequality constraints, Adaptive iterative algorithm

1. Introduction

Optimal tracking control [1], [2] is a comprehensive problem to find an optimal control input to minimize the tracking cost function to ensure that the system follows the prescribed trajectory. Due to the complexity of the nonlinear systems, it is difficult to obtain analytical solutions for OTCP. This paper aims to design an iterative algorithm to realize the tracking control without the accurate system information.

Many researchers apply Reinforcement learning (RL) technique [3], [4], [5], [6] to solve the above problems. The advantage of reinforcement learning is that it does not require a precise prior knowledge of system dynamics. Most of the methods use the ADP technique to solve the tracking Hamilton–Jacobi–Bellman (HJB) equation. Approximation methods like neural networks (NNs) are widely used in ADP literature for value function approximation. Although ADP is effective, it still has inherent technical obstacles. The HJB equation reduced to the form of an ODE under an infinite-horizon cost function and an affine nonlinear system. It is easier to learn solutions compared to PDE.

Another important problem for the OTCP is the is the existence of constraints. For control input constraints, a nonquadratic performance function [7], [8] is used in the optimal regulation problem. As for the state constraints problem, the explicit expression between optimal control and value function no longer holds in affine nonlinear systems. Therefore, within the ADP framework, the problem of state constraints for OTCP remains unresolved.

This paper aims to develop a new adaptive iterative algorithm to solve the OTCP with finite-horizon cost function and state constraints. CL technique [9] is used to learn the system identification model. Based on Pontryagin's framework and the identification model, we design a new adaptive iterative method to learn the optimal control input which minimize the tacking cost function in prescribed time interval without the exact system dynamic. Moreover, the system trajectory not only tracks the predetermined trajectory, but also satisfies the state constraints.

2. Problem statement

We consider the following nonlinear dynamic system:

$$\dot{x} = f(x,u,t),\ \ x(t_0) = x_0 \tag{1}$$

where $x \in \mathbb{R}^n$ is the system state with initial state x_0, and $u \in \mathbb{R}^m$ is the control input. The unknown map $f: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ is Lipschitz continuous. It is assumed that the state must be limited to satisfy the following inequality constraints:

$$s(x,t) \le 0 \tag{2}$$

where $s = (s_1, s_2, \dots, s_z)^T \in \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^z$ is a z-dimensional column vector function. Let $x_d(t) \in \mathbb{R}^n$ is a prescribed trajectory with initial state x_{d0}, the dynamic of tracking error and the constraints can be expressed as:

$$\begin{cases} \dot{e} = F(e,u,t) & (3) \\ S(e,t) \le 0 & (4) \end{cases}$$

where $e = x - x_d$ is the tracking error, $F(e,u,t) = f(e + x_d, u, t) - \dot{x}_d$ and $S(e,t) = s(e + x_d, t)$. We define the following finite-horizon cost function for tracking performance:

$$J(e,u) = \int_{t_0}^{t_f} L(e,u)dt \quad (5)$$

where $L(e,u) = e^T Q e + u^T R u$, $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are positive definite symmetric matrices. The optimal tracking control problem in this paper aims to find an optimal control input $u^*(t), t \in [t_0, t_f]$ to minimize the cost function (5) subject to system (3) and inequality constraints (4).

It is proved [10] that the above state constraints optimal control problem can be reduced to solve a sequence of unconstrained problem by minimizing the following Kelley-Bryson penalty cost function as $k \to \infty$:

$$J(e,u,r^k) = \int_{t_0}^{t_f} L(e,u) + r^k \sum_{j=1}^{z} h(S_j) S_j^2(e,t)\, dt \quad (6)$$

where $\lim_{k\to\infty} r^k = \infty$ and $h(\sigma) = \begin{cases} 1, \sigma > 0 \\ 0, \sigma \le 0. \end{cases}$

The necessary condition for each optimal control problem associated with $J(e,u,r^k)$ is concluded by:

$$\begin{cases} \dot{\lambda} = -H_e - 2r^k \sum_{j=1}^{z} h(S_j) S_j S_{je} \\ u = \underset{u}{\operatorname{argmin}}\, H(e,u,\lambda,t) \end{cases} \quad (7)$$

where $H(e,u,\lambda,t) = L(e,u) + \lambda^T F(e,u,t)$.

3. Concurrent learning iterative algorithm

3.1. *System identification*

First, we suppose that the unknown error dynamic (3) can be represented through a finite set of basis functions:

$$F(e,u) = \Theta^T(\mathrm{t})\Phi(e,u) + \varepsilon \quad (8)$$

where $\Phi(e,u): \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^N$ is the vector of basis functions. $\Theta: \mathbb{R} \to \mathbb{R}^{N\times n}$ is the unknown time-varying weight matrix. ε is the vector of approximation errors. According to the Weierstrass approximation theorem, ε can be reduced to zero with arbitrarily high precision as $N \to \infty$. We define the estimate of weight matrix and tracking error as $\widehat{\Theta}$ and $\hat{e}$, respectively.

For error dynamic system (3), we define the historical data stack as $\{\mathbb{E}_j(t), \mathbb{U}_j(t), \dot{\mathbb{E}}_j(t),\ t \in [t_0, t_f]\}_{j=1}^{W}$, which stores the tracking error, the control input, and the derivative of the tracking error. We have the following assumption for history trajectory stack:

Assumption 1. Let $\Pi^t = \sum_{j=1}^{W} \Phi^T(\mathbb{E}_j^t, \mathbb{U}_j^t)\, \Phi(\mathbb{E}_j^t, \mathbb{U}_j^t)$. It is assumed that there exist positive constants $\overline{\mu} > \underline{\mu} > 0$, for $\forall t[t_0, t_f]$, we have $\underline{\mu} I < \Pi^t < \overline{\mu} I$.

We design the following adaptive updating law of the i-th estimates for tracking error $\hat{e}^i$ and weight matrix $\widehat{\Theta}^i$:

$$\begin{cases} \dot{\hat{e}}^i = (\widehat{\Theta}^i(t))^T \Phi(e,u) + \omega \tilde{e}^i, \ \hat{e}^i(t_0) = x_0 - x_{d0} & (9) \\ \widehat{\Theta}^i(t) = \widehat{\Theta}^{i-1}(t) + (I - \mu\Pi^t)^2 \Phi(e,u)(\tilde{e}^i)^T \\ \quad + \mu \sum_{j=1}^{W} \Phi(\mathbb{E}_j^t, \mathbb{U}_j^t) \left(\dot{\mathbb{E}}_j^t - \left(\widehat{\Theta}^{i-1}(t)\right)^T \left(\Phi(\mathbb{E}_j^t, \mathbb{U}_j^t)\right)\right)^T \\ = \widehat{\Theta}^{i-1}(t) + (I - \mu\Pi^t)^2 \Phi(e,u)(\tilde{e}^i)^T + \mu\Pi^t \widetilde{\Theta}^{i-1}(t) \end{cases} \quad (10)$$

where $0 < \mu < \frac{1}{\overline{\mu}}$ is a positive constant. $\tilde{e}^i = e - \hat{e}^i$ and $\widetilde{\Theta}^i(t) = \Theta(\mathrm{t}) - \widehat{\Theta}^i(t)$ are i-th estimation errors of the tracking error and the weight matrix, respectively. Under Assumption 1, the adaptive updating law (10) guarantees that the weight matrix $\widehat{\Theta}^i$ converges to its true values.

3.2. *Iterative learning algorithm*

In this subsection, we will provide our iterative learning algorithm to solve the optimal tracking problem. To start with, we define:

$$\begin{cases} P(e,u,r,t) = L(e,u) + r^k \sum_{j=1}^{z} h(S_j) S_j^2(e,t) \\ Q(e,u,r,t,\lambda,\Theta) = P(e,u,r,t) + \lambda^T \Theta^T(\mathrm{t})\Phi(e,u) \end{cases} \quad (11)$$

Algorithm 1. Iterative learning algorithm for optimal tracking problem

Step 1: Initialize parameters $r^0 > 0$, $a^0 > 0$, $b^0 > 0$, $\widehat{\Theta}^0(t),\ t \in [t_0, t_f]$, convergence error $\epsilon_1 > 0$, $\epsilon_2 > 0$, $\epsilon_3 > 0$. Select the initial control input $\mathrm{u}^0(t),\ t \in [t_0, t_f]$. Update the initial estimation of tracking error by solving:

$$\dot{\hat{e}}^0 = (\widehat{\Theta}^0(t))^T\Phi(e,u) + \omega\tilde{e}^0,\ \ \hat{e}^0(t_0) = x_0 - x_{d0} \quad (12)$$

Calculate the initial tracking cost function:

$$J(e^0,u^0,r^0) = \int_{t_0}^{t_f} P(e^0,u^0,r^0,t)dt \quad (13)$$

Let $i = 0$ and $k = 0$.

Step 2: Calculate u^{i+1} and λ^i by following equation:

$$\begin{cases} \dot{\lambda}^i = -Q_e\left(e^i,u^{i+1},r^k,t,\lambda^i,\widehat{\Theta}^i\right),\ \ \lambda^i(t_f) = 0 \\ u^{i+1} = \underset{u}{argmin}\, Q\left(e^i,u,r^k,t,\lambda^i,\widehat{\Theta}^i\right) \\ \quad +a^i\left\|u-u^i\right\|_1 + \frac{1}{2}b^i\left\|u-u^i\right\|_2^2 \end{cases} \quad (14)$$

Step 3: Calculate $\hat{e}^i$ and $\widehat{\Theta}^i$ by following equation:

$$\begin{cases} \dot{\hat{e}}^i = (\widehat{\Theta}^i(t))^T\Phi(e,u) + \omega\tilde{e}^i,\ \ \hat{e}^i(t_0) = x_0 - x_{d0} \\ \widehat{\Theta}^i(t) = \widehat{\Theta}^{i-1}(t) + (I-\mu\Pi^t)^2\Phi(e,u)(\tilde{e}^i)^T \\ \quad +\mu\Pi^t\widetilde{\Theta}^{i-1}(t) \end{cases} \quad (15)$$

Step 4: Calculate the tracking cost function:

$$J(e^{i+1},u^{i+1},r^k) = \int_{t_0}^{t_f} P(e^{i+1},u^{i+1},r^k,t)dt \quad (16)$$

If $J(e^{i+1},u^{i+1},r^k) > J(e^i,u^i,r^k)$, let $a^i \leftarrow a^i + d_a$ and $b^i \leftarrow b^i + d_b$, where d_a and d_b are positive step size, then go to Step 2. Else go to Step 5.

Step 5: If $\left\|u^{i+1}-u^i\right\| \le \epsilon_1$, go to Step 6. Else let $i \leftarrow i+1$, $a^i \leftarrow a^0$, $b^i \leftarrow b^0$, go to Step 2.

Step 6: If $\left\|\widehat{\Theta}^{i+1}-\widehat{\Theta}^i\right\| \le \epsilon_2$, go to Step 7. Else let $i \leftarrow i+1$, $a^i \leftarrow a^0$, $b^i \leftarrow b^0$, go to Step 2.

Step 7: If $S(e^{i+1},t) \le \epsilon_3$, go to Step 8. Else let $i \leftarrow i+1$, $a^i \leftarrow a^0$, $b^i \leftarrow b^0$, $r^k \leftarrow r^k + d_r$ where d_r is the positive step size, then go to Step 2.

4. Simulation

In this section, we provide a nonlinear dual motor servo system with backlash to validate the efficiency of our algorithm. The system can be modeled as:

$$\begin{cases} J_i\ddot{\theta}_i + b_i\dot{\theta}_i = u_i - \tau_i \\ J_m\dot{\theta}_m + b_m\dot{\theta}_m = \sum_{i=1}^{2}\tau_i \end{cases}$$

where θ_i and θ_m represent the angles of motors and the load, J_i and J_m are the moment of inertia for the motor and the load, b_i and b_m are the resistance coefficients of the motor and load, respectively. u_i is the input torque of the motor, τ_i is the torque transmitted when the driving motor and load come into contact, which can be expressed as the following dead zone function:

$$\tau_i = \begin{cases} k(z_i+\alpha), & z_i \le \alpha \\ 0, & |z_i| < \alpha \\ k(z_i-\alpha), & z_i \ge \alpha \end{cases}$$

where k is the torque torsion coefficient, $z_i = \theta_i - \gamma\theta_m$ is the angular error between the motor and the load, γ is transmission ratio, 2α is the backlash width. Due to the non differentiability of the dead zone function, it is generally approximated by the following smooth continuous differentiable functions:

$$\tau_i = k\left(z_i - \alpha\left(\frac{2}{1+e^{-rz_i}} - 1\right)\right) = k(z_i - \alpha g_i)$$

Assuming two motors have the same parameters, the transmission ratio γ is selected as 1, the parameter r in the approximation function is selected as 10, and the parameters of the system are selected as Table 1.

Table 1. The system parameters.

Parameter	Value	Units
J_m	0.185	$kg\cdot m^2$
b_m	1.2	$N\cdot m\cdot s/rad$
J_i	0.028	$kg\cdot m^2$
b_i	1.3	$N\cdot m\cdot s/rad$
k	56	$N\cdot m/rad$
α	0.2	rad

Let $x = \left(\theta_m,\dot{\theta}_m,\theta_1,\dot{\theta}_1,\theta_2,\dot{\theta}_2\right)^T$. The target is selected as $x_d(t) = [1\ 0\ 1\ 0\ 1\ 0]^T$, we choose cost function:

$$J(e,u) = \int_0^{20}\frac{1}{2}\left(\sum_{i=1}^{6}e_i^2 + \sum_{j=1}^{2}u_j^2\right)dt$$

the state constraints are assumed to be $|x_2| \le 0.4rad/s$, $|x_4| \le 0.4rad/s$, $|x_6| \le 0.4rad/s$, which can be rewritten as: $s_1 = e_2 - 0.4$, $s_2 = -e_2 - 0.4$, $s_3 = e_4 - 0.4$, $s_4 = -e_4 - 0.4$, $s_5 = e_6 - 0.4$, $s_6 = -e_6 - 0.4$. The basis function is selected as $\Phi(e,u) = (e_1,e_2,e_3,e_4,e_5,e_6,g_1(e_1,e_3),g_2(e_1,e_5),u_1,u_2)^T$. The initial state is $x_0 = (1.5\ \ 0\ \ 1.6\ \ -0.1\ \ 1.5\ \ 0.1)^T$. The initial parameters are set to be $r_0 = c_0 = d_0 = 1$, $\widehat{\Theta}_0^0 = rand(6,10)^T$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = 0.001$.

Fig. 1 and Fig. 2 are visual representations of the results. We can see that compared with the initial trajectory, the angles of load and motor have stabilized around the predetermined values.

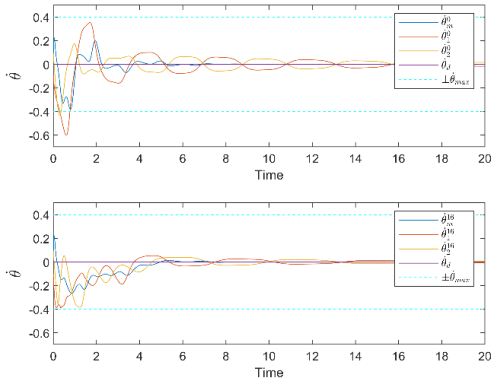

Fig. 1. Comparisons of the tracking angles.

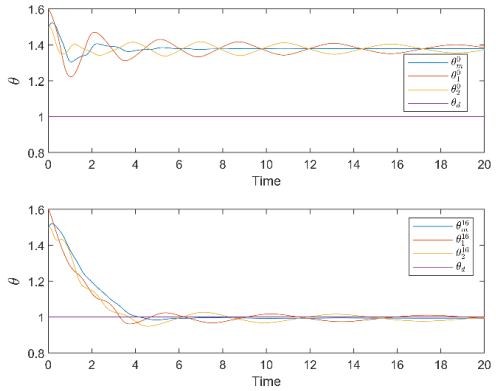

Fig. 2. Comparisons of the tracking angular velocities.

5. Conclusion

In this paper, we present a novel concurrent iterative learning algorithm. The tracking cost function is selected as a finite-horizon form. Our method does not require precise dynamic information and can handle with the time-varying systems. Moreover, state constraints are considered in the process of solving OTCP.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61973044).

References

1. Firdaus E Udwadia. Optimal tracking control of nonlinear dynamical systems. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 464(2097):2341–2363, 2008.
2. Mohamed Boukattaya, Mohamed Jallouli, and Tarak Damak. On trajectory tracking control for nonholonomic mobile manipulators with dynamic uncertainties and external torque disturbances. Robotics and autonomous systems, 60(12):1640–1647, 2012.
3. Hamidreza Modares and Frank L Lewis. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. Automatica, 50(7):1780–1792, 2014.
4. Ruizhuo Song, Frank L Lewis, Qinglai Wei, and Huaguang Zhang. Off policy actor-critic structure for optimal control of unknown systems with disturbances. IEEE transactions on cybernetics, 46(5):1041–1050, 2015.
5. Yuanheng Zhu, Dongbin Zhao, and Xiangjun Li. Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics. IET Control Theory & Applications, 10(12):1339–1347, 2016.
6. Fayez El-Sousy, Mahmoud M Amin, and Ahmed Al-Durra. Adaptive optimal tracking control via actor-critic-identifier based adaptive dynamic programming for permanent-magnet synchronous motor drive system. IEEE Transactions on Industry Applications, 57(6):6577–6591, 2021.
7. Murad Abu-Khalaf and Frank L Lewis. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach. Automatica, 41(5):779–791, 2005.
8. Bahare Kiumarsi, Frank L Lewis, Hamidreza Modares, Ali Karimpour, and Mohammad-Bagher Naghibi-Sistani. Reinforcement q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics. Automatica, 50(4):1167–1175, 2014.
9. Girish V Chowdhary and Eric N Johnson. Theory and flight-test validation of a concurrent-learning adaptive controller. Journal of Guidance, Control, and Dynamics, 34(2):592–607, 2011.
10. Milind M Lele and David H Jacobson. A proof of the convergence of the kelley-bryson penalty function technique for state-constrained control problems. Journal of Mathematical Analysis and Applications, 26(1):163–169, 1969.

Authors Introduction

Mr. Yuqi Zhang

He received the B.S. degree in Automation from Beijing University of Posts and Telecommunications, Beijing, China, in 2020. He is currently working toward the M.D. degree in Control Science and Engineering. His research interests include differential games and reinforcement learning.

Dr. Bin Zhang

He received his B.S. and Ph.D. degrees both in control theory and applications from Beihang University, Beijing, China, in 2010 and 2016, respectively. He is currently an Associate Professor with the School of Artificial Intelligence at Beijing University of Posts and Telecommunications. His research interests include reinforcement learning, multi-agent systems, and intelligent control.