EasyChair Preprint

№ 3455

Linear Quadratic Tracking Control of Partial Unknown Continuous-Time Systems Using Integral Reinforcement Learning

Cheng Weiran, Xiao Zhenfei and Li Jinna

May 21, 2020

Linear Quadratic Tracking Control of Partial Unknown Continuous-Time Systems Using Integral Reinforcement Learning

Weiran Cheng Zhenfei Xiao Jinna Li*, *Member, IEEE,*

***Abstract*—In this paper, an integral reinforcement learning (IRL) algorithm is proposed for solving the linear quadratic tracking (LQT) problem of partial unknown continuous-time systems that try to chase a polynomial reference signal. By using IRL technique to solve the algebraic Riccati equation (ARE) derived from LQT problem, the approximate optimal tracking controller can be obtained without fully understanding the system drift dynamics and command generator dynamics. Firstly, the optimal tracking control problem is formulated. The augmented vector is defined, and the algebraic Riccati equation is obtained based on the dynamic programming method. Then, employing IRL yields the iterative Bellman equation and policy updating expression, such that an IRL algorithm is finally developed for finding the approximate optimal tracking control policy, under which the reference signal with higher-order polynomials and unknown model parameters can be tracked by a linear system with partial known model parameters, meanwhile the specific performance can be minimized. Finally, a simulation example is given to verify the efficiency of the provided mathod.**

***Index Terms*—Linear quadratic tracking, policy iteration, integral reinforcement learning.**

I. Introduction

THE optimal tracking problem which is called the linear quadratic tracking (LQT) is an important problem for control systems design. The goal of LQT is to design a controller such that the output signal of the system can track a specific reference signal optimally by minimizing a performance indicator. The traditional solution to solve LQT problem to compute the feedback term by using an algebraic Riccati equation (ARE) and the feedforward term based on a noncausal difference equation [1].

Reinforcement learning (RL), as a kind of machine learning methods, has been widely employed to learn the optimal controllers of the systems by using complete known dynamics of systems [2–5]. However, for many practical problems, it is difficult to obtain the system dynamics. Scholars have concentrated on developing RL techniques that derive optimal controllers for unknown dynamical systems. For continuous-time (CT) dynamical systems, RL techniques were first employed by Werbos to seek solutions to the optimal regulator problem for discrete-time systems [5]. Doya presented RL framework without a prior discretization of time, state and control [6]. RL method also used to solve H∞ control problem, such as the Model-free Q-learning method [7] and the neural dynamic programming [8]. RL technology also used to solve the optimal tracking problem of discrete-time (DT) system and solve the LQT problem of DT systems without requiring the knowledge of the systems, such as application of iterative adaptive dynamic programming (ADP) and greedy heuristic dynamic programming (HDP) iteration algorithm in nonlinear DT systems [9, 10], a class of nonlinear discrete-time systems with time delays based on HDP [11], and policy iteration (PI) for discrete-time linear systems [12]. The existing LQT results using RL methods usually can work for following a constant or linear dynamical reference signal [13–20]. But few scholars used RL algorithms to solve the LQT problem for CT systems with the objective of tracking an unknown polynomial reference signal [13].

In this paper, based on RL, an adaptive controller is developed to solve the LQT problem of unknown CT linear systems. Firstly, assume that the reference trajectory is generated by a command generator system which is a custom polynomial function. By setting the difference between the system trajectory and the reference trajectory, the augmentation system is established on the basis of the original system, and the value function of LQT problem is transformed into a quadratic structure. Then based on the value function, a new Bellman equation is derived. Secondly, we proposed the IRL algorithm to learn the solution to the LQT problem using partial knowledge about the system dynamics. The convergence of the proposed algorithm to the optimal control solution is verified by simulation, and the effectiveness of the proposed method is proved.

II. Continuous-time linear quadratic tracking problem formulation

In this section, the infinite LQT problem is formulated for CT systems.

Consider the CT linear system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{1}$$

This work was supported by the National Natural Science Foundation of China under Grants 61673280, the Open Project of Key Field Alliance of Liaoning Province under Grant 2019-KF-03-06 and the Project of Liaoning Shihua. University under Grant 2018XJJ-005.

Weiran Cheng is with the School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, P.R. China. (18341315179@163.com)

Zhenfei Xiao is with the School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, P.R. China. (Xiaozhenfeiwm@outlook.com)

Jinna Li is with the School of Information and Control Engineering, Liaoning Shihua University, Liaoning 113001, P.R. China. Jinna Li is the corresponding author (lijinna_721@126.com)

where x is a measurable system state vector , u is the control input, y is the output and A, B, C are matrices with compatible dimensions. Find an optimal control policy u^* which makes the output y track a desired trajectory y_r.

The performance index is usually formulated as

$$J(x, y_r) = \int_t^\infty \left[(Cx - y_r)^T Q (Cx - y_r) + u^T R u \right] d\tau \quad (2)$$

where $Q > 0$ and $R > 0$ are symmetric matrices.

III. Derivation of solving LQT problem

In this section, we propose a solution to the LQT problem. Suppose the reference trajectory is generated by a command generator system which is a customized polynomial function. By setting the difference between the system trajectory and the reference trajectory, the augmentation system is established on the basis of the original system, and the value function of LQT problem is transformed into a quadratic structure. Using the quadratic structure of the value function, a novel Bellman equation is derived for the LQT problem.

Assumption 1: reference trajectory is generated by a polynomial function

$$y_r = a_0 + a_1 t + a_2 t^2 + \cdots + a_{d-1} t^{d-1} \quad (3)$$

where a_0, a_1, a_2 ...., a_{d-1} are matrices with appropriate dimensions. Let $e = y - y_r$, suppose $\dot{q}_1 = e$, $\dot{q}_2 = q_1$ .... , $\dot{q}_d = q_{d-1}$ then

$$q_d{}^{(d)} = e \quad (4)$$

Define the new augmented system state as

$$\dot{z}(t) = [\dot{x}(t) \;\; \dot{q}_1 \;\; \cdots \;\; \dot{q}_d] \quad (5)$$

$$z(t) = [x(t) \;\; q_1 \;\; \cdots \;\; q_d] \quad (6)$$

Putting (1) and (2) together to construct the augmented system as

$$\dot{z}(t) = A_z z(t) + B_z u(t) + M y_r \quad (7)$$

where

$$A_z = \begin{bmatrix} A & 0 & \cdot & \cdot & \cdot & 0 \\ C & 0 & & & & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix} \quad (8)$$

$$B_z = [B \;\; 0 \;\; \cdots \;\; 0] \quad (9)$$

$$M = [0 \;\; -1 \;\; 0 \;\; \cdots \;\; 0] \quad (10)$$

Define the performance index

$$\begin{aligned} J^* &= \min_{u(t)} z(t)^T Q z(t) + u(t)^T R u(t) \\ s.t \quad & \dot{z}(t) = A_z z(t) + B_z u(t) \end{aligned} \quad (11)$$

Then define the value function

$$V(z(t)) = \int_t^\infty \left[z^T(\tau) Q z(\tau) + u^T(\tau) R u(\tau) \right] d\tau \quad (12)$$

For the optimal tracking problem, the goal is to find the control policy $u(t)$ depending on $z(t)$. Consider the control policy

$$u(t) = K z(t) \quad (13)$$

The value function (12) with control policy (13) can be written as the quadratic form

$$V(z(t)) = z^T(t) P z(t) \quad (14)$$

where $P > 0$ is symmetric matrices

Define the Hamiltonian

$$H\left(z(t), \frac{\partial v}{\partial t}, u\right) = \dot{V} + z^T(t) Q z(t) + u^T(t) R u(t) \quad (15)$$

According to Hamilton function (15) and augmented system (7), the augmented LQT Bellman equation is given as

$$0 = \dot{z}(t)^T P z(t) + z(t)^T P \dot{z}(t) + z^T Q z + u^T R u \quad (16)$$

The optimal control solution for the LQT problem is given by

$$u^* = K z(t) \quad (17)$$

where

$$K = -R^{-1} B_z^T P z(t) \quad (18)$$

and P satisfies the LQT Riccati equation (ARE)

$$0 = A_z{}^T P + P A_z - P B_z R^{-1} B_z{}^T P + Q \quad (19)$$

Remark 1: Based on [21], the control policy (18) can render the output of system (1) to follow the reference trajectory (3).

IV. Integral reinforcement learning for solving the LQT

In this section, we first use an offline PI Algorithm to solve the LQT problem. Then, employing integral reinforcement learning yields the iterative Bellman equation and policy updating expression, such that an IRL algorithm is finally developed for finding the approximate optimal tracking control policy of the LQT problem using partial knowledge about the system dynamics.

A. Offline Policy Iteration Algorithm for Solving The LQT ARE

Algorithm 1 Offline policy iteration for solving the LQT problem

1. Initiation: given $u^0 = K^0 z(t)$, which is admissible;
2. Policy evaluation: Find P^i using the LQT Lyapunov equation

$$\left(A_z + B_z K^i\right)^T P^i + P^i \left(A_z + B_z K^i\right) + \left(K^i\right)^T R \left(K^i\right) + Q = 0$$

3. Policy updating: update the control gain using

$$K^{i+1} = -R^{-1} B_z^T P^i$$

B. IRL Algorithm For Solving LQT Problem

Suppose time interval $T > 0$ to obtain the IRL Bellman equation, the value function (13) satisfies

$$V(z(t)) = \int_t^{t+T} (z^T(\tau)Qz(\tau) + (u(\tau))^T Ru(\tau))d\tau + V(z(t+T)) \tag{20}$$

The LQT IRL Bellman equation can be obtained by using (14) in (20)

$$z^T(t)Pzt(t) = \int_t^{t+T} (z^T(\tau)Qz(\tau) + (u(\tau))^T Ru(\tau))d\tau + z^T(t+T)Pz(t+T) \tag{21}$$

Algorithm 2 Online IRL algorithm for solving the LQT problem

1. Initiation: Given initial admissible control gain K^i, and let $i = 0$, where i denotes iteration index, we have $u^0 = K^0 z(t)$;
2. Policy evaluation: Given a control policy u^i, find P^i using the Bellman equation

$$z^T(t)P^i z(t) = \int_t^{t+T} (z^T(\tau)Qz(\tau) + (u^i(\tau))^T Ru^i(\tau))d\tau + z^T(t+T)P^i z(t+T)$$

3. Policy updating: update the control gain using

$$u^{i+1} = -R^{-1}B_z^T P^i z(t)$$

4. Stop when $||P^{i+1} - P^i|| \leq \varepsilon$ with a small constant ε ($\varepsilon > 0$).

V. Simulation Results

In this section, a simulation example is given to verify the effective of Algorithm 2 for solving the LQT problem.

Consider the continuous-time linear system:

$$\dot{x}(t) = \begin{bmatrix} 0.5 & 1.5 \\ 2.0 & -2 \end{bmatrix} x(t) + \begin{bmatrix} 5 \\ 1 \end{bmatrix} u(t)$$
$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) \tag{22}$$

and suppose that the state trajectory is generated by a polynomial function

$$y_r = a_0 \tag{23}$$

with $a_0 = 10$. The performance index is given as (2), choose $Q = [5\ 0\ 5;\ 0\ 0\ 0;\ 5\ 0\ 5]$ and $R = 1$. The solution can be obtained in terms of the LQT ARE (19) using the command 'dare' in Matlab, i.e.

$$P^* = \begin{bmatrix} 0..5034 & 0.0529 & 0.4441 \\ 0.0529 & 0.0195 & 0.0154 \\ 0.4441 & 0.0154 & 0.4931 \end{bmatrix} \tag{24}$$

and the optimal control gain K^* can be obtained by using (18)

$$K^* = \begin{bmatrix} -2.5697 & -0.2841 & -2.2361 \end{bmatrix} \tag{25}$$

Implementing Algorithm 2 to solve the LQT problem of the system. Fig. 1 shows the convergence evolutions of the optimal P matrix. Also, Fig. 2 depicts the convergence evolutions of the optimal control gain. From Figs. 1 and 2, it is clear that the value function matrix and the control gain converge to their optimal values after four iterations. Fig. 3 shows that the output finally tracks the state trajectory of the system.

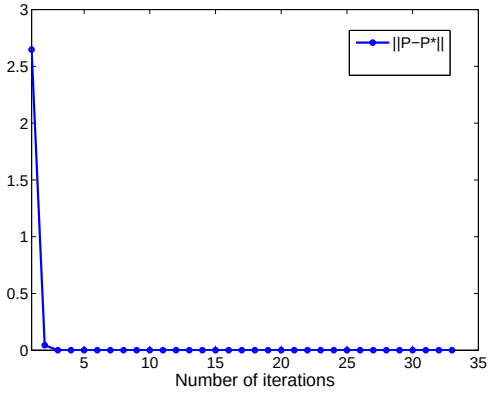

Fig. 1. Convergence of the P matrix to its optimal value when d=1

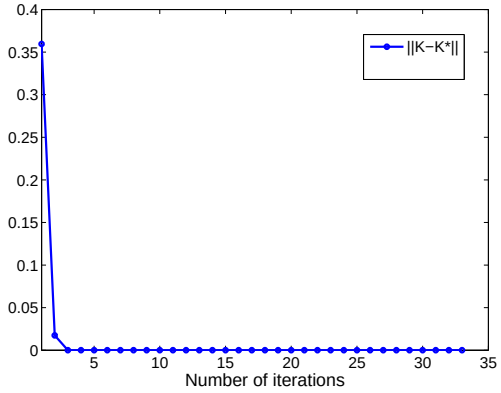

Fig. 2. Convergence of the controller gain to its optimal value when d=1

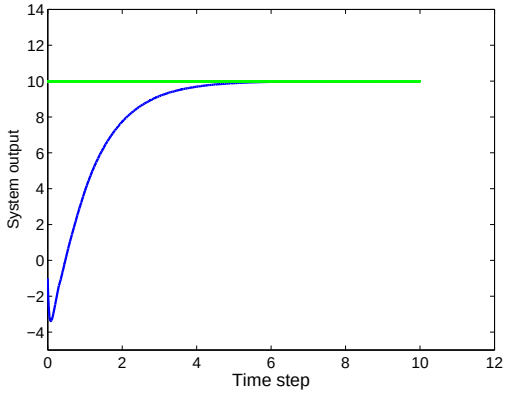

Fig. 3. System output versus reference trajectory when d=1

When $d = 2$, assume that the desired trajectory is generated by the command generator system

$$y_r = a_0 + a_1 t \tag{26}$$

with $a_0 = 1$, $a_1 = 2$. The performance index is given as (2), choose $Q = [5\ 0\ 0\ 5;\ 0\ 0\ 0\ 0;\ 0\ 0\ 0\ 0;\ 5\ 0\ 0\ 5]$ and $R = 1$. The solution can be obtained in terms of the LQT ARE (19) using the command 'dare' in Matlab, i.e.

$$P^* = \begin{bmatrix} 0.4764 & 0.0525 & 0.1199 & 0.4438 \\ 0.0525 & 0.0195 & 0.0123 & 0.0173 \\ 0.1199 & 0.0123 & 0.9605 & 0.1870 \\ 0.4438 & 0.0173 & 0.1870 & 1.3674 \end{bmatrix} \tag{27}$$

and the optimal control gain K^* can be obtained by using (18)

$$K^* = \begin{bmatrix} -2.4344 & -0.2823 & -0.6115 & -2.2361 \end{bmatrix} \tag{28}$$

Implementing Algorithm 2 to solve the LQT problem of the system. Fig. 4 shows the convergence evolutions of the optimal P matrix. Also, Fig. 5 depicts the convergence evolutions of the optimal control gain. From Figs. 4 and 5, it is clear that the value function matrix and the control gain converge to their optimal values. Fig. 6 shows that the output finally tracks the state trajectory of the system.

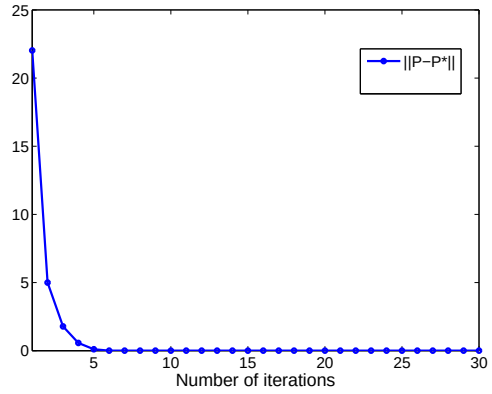

Fig. 4. Convergence of the P matrix to its optimal value when d=2

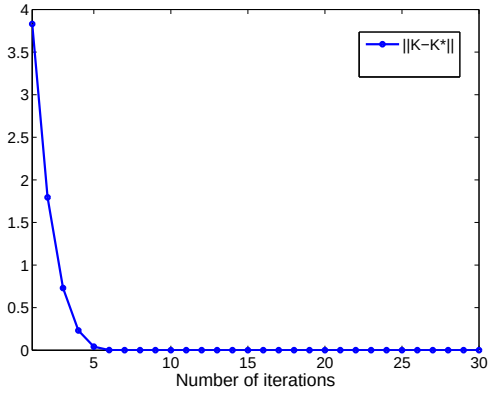

Fig. 5. Convergence of the controller gain to its optimal value when d=2

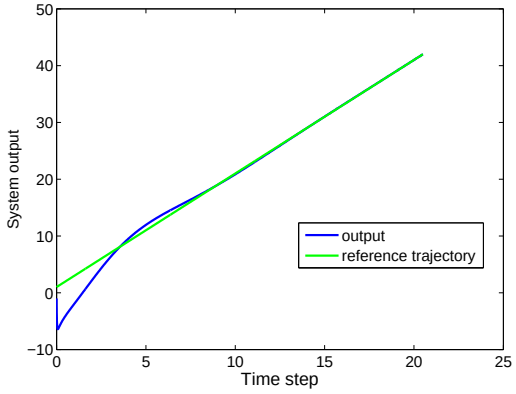

Fig. 6. System output versus reference trajectory when d=2

VI. Conclusion

An IRL algorithm is proposed to solve the LQT problem for partially-unknown CT linear systems with the objective of tracking a polynomial reference signal. On the basis of the value function which has quadratic form in terms of the state and the reference trajectory, a LQT ARE was obtained and a PI based IRL algorithm is developed to solve the LQT ARE online without requiring full knowledge of system dynamics and with no need of model parameters of reference signal. The simulation results have shown that the proposed formulation for the LQT problem gives a satisfactory tracking performance.

References

[1] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control, 3rd Edition*. Wiley, New York, 2012.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning‡An Introduction*. MA: MIT Press, Cambridge, 1998.

[3] J. Li, H. Modares, T. Chai, F. L. Lewis, and L. Xie, "Off-policy reinforcement learning for synchronization in multi-agent graphical games," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2434–2445, 2017.

[4] B. Kiumarsi, F. L. Lewis, and Z. Jiang, "H∞ control of linear discrete-time systems: Off-policy reinforcement learning," *Automatica*, pp. 144–152, 2017.

[5] W. B. Powell, "Approximate dynamic programming: Solving the curses of dimensionality," 2007.

[6] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, no. 1, pp. 219–245, 2000.

[7] J. Li, B. Kiumarsi, T. Chai, F. L. Lewis, and J. Fan, "Off-policy reinforcement learning: optimal operational control for two-time-scale industrial processes," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4547–4558, 2017.

[8] M. Abukhalaf, F. L. Lewis, and J. Huang, "Neurodynamic programming and zero-sum games for constrained control systems," *IEEE Transactions on Neural Networks*, vol. 19, no. 7, pp. 1243–1252, 2008.

[9] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration

algorithm," *IEEE Transactions on Systems Man & Cybernetics Part B*, vol. 38, no. 4, pp. 937–942, 2008.

[10] D. R. Wei Q. L., Liu, "Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 4, pp. 1020–1036, 2014.

[11] H. Zhang, R. Song, Q. Wei, and T. Zhang, "Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 1851–1862, 2011.

[12] B. Kiumarsi, F. L. Lewis, M. B. Naghibi-Sistani, and A. Karimpour, "Optimal tracking control of unknown discrete-time linear systems using input-output measured data," *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2770–2779, 2015.

[13] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226–2236, 2011.

[14] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm," *IEEE Transactions on Systems Man & Cybernetics Part B*, vol. 38, no. 4, pp. 937–942, 2008.

[15] T. Dierks and S. Jagannathan, "Optimal tracking control of affine nonlinear discrete-time systems with unknown internal dynamics," in *IEEE Conference on Decision & Control*, 2010.

[16] Q. Wei and D. Liu, "Optimal tracking control scheme for discrete-time nonlinear systems with approximation errors," in *International Symposium on Neural Networks*, 2013.

[17] J. Li, J. Ding, T. Chai, and F. L. Lewis, "Nonzero-sum game reinforcement learning for performance optimization in large-scale industrial processes," *IEEE Transactions on Cybernetics*, doi:10.1109/TCYB.2019.2950262.

[18] Y. Huang and D. Liu, "Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative adp algorithm," *Neurocomputing*, vol. 125, pp. 46–56, 2014.

[19] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.

[20] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks the Official Journal of the International Neural Network Society*, vol. 22, no. 3, pp. 237–246, 2009.

[21] H. Tan, S. Shu, and F. Lin, "An optimal control approach to robust tracking of linear systems," *International Journal of Control*, vol. 82, no. 3, pp. 525–540, 2009.