Distributed Q-Learning with State Tracking for Multi-agent Networked Control

Extended Abstract

Hang Wang
Arizona State University
Tempe, Arizona
hwang442@asu.edu

Sen Lin
Arizona State University
Tempe, Arizona
slin70@asu.edu

Hamid Jafarkhani
University of California, Irvine
Irvine, California
hamidj@uci.edu

Junshan Zhang
Arizona State University
Tempe, Arizona
Junshan.Zhang@asu.edu

ABSTRACT

This paper studies distributed Q-learning for Linear Quadratic Regulator (LQR) in a multi-agent network. The existing results often assume that agents can observe the global system state, which may be infeasible in large-scale systems due to privacy concerns or communication constraints. In this work, we consider a setting with unknown system models and no centralized coordinator. We devise a state tracking (ST) based Q-learning algorithm to design optimal controllers for agents. Specifically, we assume that agents maintain local estimates of the global state based on their local information and communications with neighbors. At each step, every agent updates its local global state estimation, based on which it solves an approximate Q-factor locally through policy iteration. Assuming a decaying injected excitation noise during the policy evaluation, we prove that the local estimation converges to the true global state, and establish the convergence of the proposed distributed ST-based Q-learning algorithm. The experimental studies corroborate our theoretical results by showing that our proposed method achieves comparable performance with the centralized case.

KEYWORDS

Reinforcement Learning; Multi-agent; Linear Quadratic Control

ACM Reference Format:
Hang Wang, Sen Lin, Hamid Jafarkhani, and Junshan Zhang. 2021. Distributed Q-Learning with State Tracking for Multi-agent Networked Control: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

One main objective in the distributed control of multi-agent systems (MASs) is to learn local controllers for agents in a distributed manner so as to minimize the global cost. The design of distributed controllers is challenging due to the networked nature of MASs. Observe that the agents are physically coupled with certain interconnections [5], e.g., the buses in a microgrid are interconnected through structural links such as the power transmission lines. Consequently, the controller synthesis at a bus has to account for the impact of other buses. To deal with the sophisticated coupling in MASs, the model-based distributed controller design has been studied in [1, 4, 7]. These studies assume that the underlying system model is *known*, which may be infeasible in large-scale systems.

Notably, data-driven Q-learning [10], which is a model-free Reinforcement Learning (RL) approach [2], has been proposed to learn the optimal LQR controller online in the single agent case [3]. Most recent works apply the Q-learning in the multi-agent LQR control and show that good performance can be achieved assuming that the knowledge of global state information is shared by a centralized coordinator [6, 8]. Nevertheless, such a centralized coordinator is often not available in many scenarios.

In this work, we consider distributed LQR control in MASs with only partial observations. We propose a novel distributed Q-learning approach with state tracking (ST-Q), where each agent first constructs a global state estimator based on local communication with its neighbors, and then solves an approximate Q-learning problem accordingly. Intuitively, by exchanging state estimations with neighboring agents, an individual agent would be able to improve its global state estimator as the information continuously diffuses across the network. The convergence of distributed Q-learning algorithms in multi-agent LQR control has been underexplored. In this work, we fill this void and establish the convergence of the proposed distributed ST-Q algorithm. Our proposed method achieves comparable performance with the the full observation case [8].

2 STATE TRACKING METHODS

Distributed LQR Control. Consider a multi-agent network consisting of L agents, where the Linear Time Invariant (LTI) system dynamics at each agent $i \in [L]$ is given as follows:

$$x_i(t+1) = \textstyle\sum_{j=1}^{L} A_{ij} x_j(t) + B_i u_i(t) \qquad (1)$$

where $x_i(t) \in \mathbb{R}^n$ is Agent i's state vector and $u_i(t) \in \mathbb{R}^m$ is its control input at time t. A_{ij} and B_i are unknown system parameters. For the subsystem (1) at each agent i, the stage cost incurred by executing the control $u_i(t)$ in state $x_i(t)$ at time t is given by $g_i(x_i(t), u_i(t)) = x_i(t)^\top P_i x_i(t) + u_i(t)^\top R_i u_i(t)$, where P_i and R_i are

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

Algorithm 1 ST based Q-learning (ST-Q)

```
Require: $K_{i1}$: initial stable controller, $\theta_{i1}(0) = 0$: initial estimation,
    $q = 1$: policy improvement index, $\varepsilon_K$: tolerance error
repeat
    for Agent $i = 1, \cdots, L$ do
        Estimate global state using State Tracking (9)
        Estimate $\theta_{iq}$ by solving (8) (e.g., SGD).
    end for
    for Agent $i = 1, \cdots, L$ do
        Update policy $K_{i(q+1)} = -H_{iq,22}^{-1} H_{iq,21}$ using $\hat{\theta}_{iq}$.
        Set $\hat{\theta}_{i(q+1)}(0) = \hat{\theta}_{iq}$.
    end for
    Set $q = q + 1$.
until $\|\hat{\theta}_{i(q+1)} - \hat{\theta}_{iq}\| < \varepsilon_K, \forall i \in [L]$
```

positive semi-definite matrices. Let $J_i(x_i(0)) = \sum_{\tau=0}^{\infty} g_i(x_i(\tau), u_i(\tau))$ denote the local cost function at Agent i.

Let $x_{\mathcal{N}_i}(t) \in \mathcal{X}_{\mathcal{N}_i}$ denote the state information available for Agent i at time t, which contains partial entries of the global state vectors from its neighbors $\mathcal{N}_i$ in the graph. Agent i then selects the local control input $u_i(t) \in \mathcal{U}_i$, based on the information $x_{\mathcal{N}_i}(t)$ and a control policy $\tilde{\pi}_i$ with a linear feedback controller, i.e.,

$$\tilde{\pi}_i : \mathcal{X}_{\mathcal{N}_i} \mapsto \mathcal{U}_i. \tag{2}$$

We further assume that K_i is the feedback controller in the policy $\tilde{\pi}_i$. The goal of distributed LQR control with local communication is to find controllers that minimize the global cost function $J(X(0))$:

$$\min_{\{K_i\}} \; J(X(0)) = \textstyle\sum_{i=1}^{L} J_i(x_i(0)), \text{ s.t. (1), (2).} \tag{3}$$

In this work, we aim to achieve the optimal controller K_i^* for each agent i that is the same as in the case where the model parameters are known, by solving Problem (3).

Q-learning and Policy Iteration. Given the global state $X(t)$ and the local control policy $\pi_i : \; u_i(t) = K_i X(t)$ for some state feedback controller K_i, Q-learning defines the Q-factor for agent i as follows:

$$Q_i(x_i(t), u_i(t)) = g_i(x_i(t), u_i(t)) + J_i(x_i(t+1)). \tag{4}$$

It can be shown that (4) can be rewritten as the quadratic from:

$$Q_i(x_i(t), u_i(t)) = [X(t); u_i(t)]^\top H_i [X(t); u_i(t)], \tag{5}$$

where $H_i = [H_{i,11}, H_{i,12}; H_{i,21}, H_{i,22}]$ is a symmetric block matrix [2]. Suppose we have determined the Q-factor Q_i for a controller K_i. The policy improvement step aims to find a better controller:

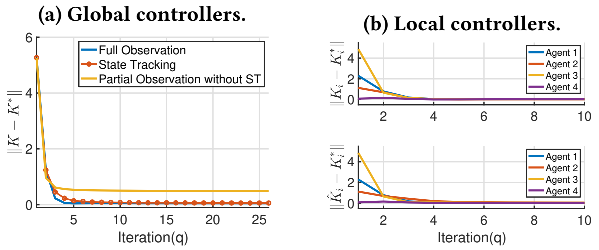
$$K_i^{\text{new}} = \arg\min_{K_i} (Q_i(x_i(t), K_i X(t))) = -H_{i,22}^{-1} H_{i,21}. \tag{6}$$

To determine the matrix H_i in the policy evaluation step, along the same line as in [3], we reformulate the quadratic form of $Q_i(x_i(t), u_i(t))$ in (5) in a linear form parameterized by θ_i:

$$Q_i(x_i(t), u_i(t)) = y_i(t)^\top \theta_i, \tag{7}$$

where $y_i(t) = [x_1^2(t), x_1(t)x_2(t), \cdots, x_L(t)u_i(t), u_i^2(t)]$ is a vector containing all of the quadratic basis over the elements in $[X(t); u_i(t)]$, and the parameter θ_i is obtained through some manipulation after removing the redundant elements of the symmetric matrix H_i.

Based on the linear form (7), by observing sufficient samples of the stage cost $g_i(x_i(t), u_i(t))$ and $\phi_i(t)$, θ_i can be obtained by solving a least square estimation problem (Policy Evaluation):


(a) Global controllers.

(b) Local controllers.

Figure 1: Convergence comparisons among three cases.

$$g_i(x_i(t), u_i(t)) = (y_i(t) - y_i(t+1))^\top \theta_i \triangleq \phi_i(t)^\top \theta_i. \tag{8}$$

State Tracking. To address the issue that the global state $X(t)$ is not available, we propose a state tracking scheme to facilitate the estimation of the global state $X(t)$ through the information exchange among agents. At time t each agent i maintains a local estimation $Z_i(t)$ of the global state $X(t)$:

$$Z_i(t) = \operatorname{col}(\bar{x}_{i1}(t), \bar{x}_{i2}(t), \cdots, \bar{x}_{iL}(t)),$$

where $\bar{x}_{ij}(t)$ is the estimation of Agent j's state $x_j(t)$ at Agent i for time t. In particular, $\bar{x}_{ii}(t) = x_i(t)$. At time $t+1$, each agent i first receives the state $x_j(t+1)$ from every neighbors, and then updates the corresponding entries in its estimation $Z_i(t)$, i.e., $\bar{x}_{ij}(t) \longleftarrow x_j(t+1)$, $\forall j \in \mathcal{N}_i$. Consequently, an updated estimation $\hat{Z}_i(t+1) = \operatorname{col}(\hat{x}_{i1}(t+1), \hat{x}_{i2}(t+1), \cdots, \hat{x}_{iL}(t+1))$ can be obtained with

$$\hat{x}_{ij}(t+1) = \begin{cases} \bar{x}_{ij}(t) & \forall j \notin \mathcal{N}_i, \\ x_j(t+1) & \forall j \in \mathcal{N}_i. \end{cases}$$

Next, each agent i shares its updated global state estimation $\hat{Z}_i(t+1)$ with its neighbors. After receiving the global state estimation $\hat{Z}_i(t+1)$ from the neighboring agents, Agent i computes the state estimation $\bar{x}_{ij}(t+1)$ by taking a weighted average of the corresponding estimations $\hat{x}_{kj}(t+1)$ from its neighbors $k \in \mathcal{N}_i$. The weighting process is modeled by a doubly stochastic weight matrix, $W = [w_{ij}]$. The specific update rule is shown as following

$$\bar{x}_{ij}(t+1) = \begin{cases} \sum_{k=1}^{L} w_{ik} \hat{x}_{kj}(t+1) & \forall j \notin \mathcal{N}_i, \\ x_j(t+1) & \forall j \in \mathcal{N}_i. \end{cases} \tag{9}$$

3 RESULTS AND CONCLUSION

In the technical report [9], we theoretically analyze the convergence of the proposed ST-Q learning algorithm (Algorithm 1) under mild assumptions. Empirically, we demonstrate the performace of ST-Q learning on a four-agent LTI system with unknown dynamics. As shown in Fig. 1a, the controller obtained by the ST-Q learning approach eventually converges to the optimal controller. Moreover, Fig. 1b further demonstrates the convergence performance of the local controller $\hat{K}_i$ at each agent i compared with distributed Q-learning with global state (DQG), i.e., each agent in the ST-Q learning almost has the same convergence behaviour as in DQG.

ACKNOWLEDGMENTS

This work is supported in part by NSF Grants CNS-2003081 and CPS-1739344.

REFERENCES

[1] Gianluca Antonelli. 2013. Interconnected dynamic systems: An overview on distributed control. *IEEE Control Systems Magazine* 33, 1 (2013), 76–88.

[2] Dimitri P Bertsekas. 1995. *Dynamic programming and optimal control*. Vol. 1. Athena scientific Belmont, MA.

[3] Steven J Bradtke, B Erik Ydstie, and Andrew G Barto. 1994. Adaptive linear quadratic control using policy iteration. In *Proceedings of 1994 American Control Conference-ACC'94*, Vol. 3. IEEE, 3475–3479.

[4] Christian Conte, Colin N Jones, Manfred Morari, and Melanie N Zeilinger. 2016. Distributed synthesis and stability of cooperative distributed model predictive control for linear systems. *Automatica* 69 (2016), 117–125.

[5] Derui Ding, Qing-Long Han, Zidong Wang, and Xiaohua Ge. 2019. A survey on model-based distributed control and filtering for industrial cyber-physical systems. *IEEE Transactions on Industrial Informatics* 15, 5 (2019), 2483–2499.

[6] Amirhassan Fallah Dizche, Aranya Chakrabortty, and Alexandra Duel-Hallen. 2019. Sparse wide-area control of power systems using data-driven reinforcement learning. In *2019 American Control Conference (ACC)*. IEEE, 2867–2872.

[7] Paolo Massioni and Michel Verhaegen. 2009. Distributed control for identical dynamically coupled systems: A decomposition approach. *IEEE Trans. Automat. Control* 54, 1 (2009), 124–135.

[8] Vignesh Narayanan and Sarangapani Jagannathan. 2016. Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid Q-learning approach. *IET Control Theory & Applications* 10, 12 (2016), 1448–1457.

[9] Hang Wang, Sen Lin, Hamid Jafarkhani, and Junshan Zhang. 2020. Distributed Q-Learning with State Tracking for Multi-agent Networked Control. *arXiv preprint arXiv:2012.12383* (2020).

[10] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3-4 (1992), 279–292.