Model Reference Adaptive Control for Online Policy Adaptation and Network Synchronization

Miguel F. Arevalo-Castiblanco [1] **César A. Uribe** [2] **Eduardo Mojica-Nava** [1]

Abstract

We propose an online adaptive synchronization method for leader-follower networks of heterogeneous agents. Synchronization is achieved using a distributed Model Reference Adaptive Control (DMRAC-RL) that enables the improved performance of Reinforcement Learning (RL)-trained policies on a reference model. The leader observes the performance of the reference model, and the followers observe the states and actions of the agents they are connected to, but not the reference model. Notably, both the leader and followers models might differ from the reference model the RL-control policy was trained. DMRAC-RL uses an internal loop that adjusts the learned policy for the agents in the form of an augmented input to solve the distributed control problem. Numerical examples of the synchronization of a network of inverted pendulums support our theoretical findings.

1. Introduction

The outstanding interest in cooperative control of multiagent systems (MAS) stems from current and future critical applications such as autonomous multi-vehicles systems, resource allocation in networks, synchronization in power systems, and many more (Lewis et al., 2013; Wang et al., 2017). Consensus-based control strategies have been pervasive in developing cooperative control in MAS, and extensive literature is available beginning with the seminar works (Bertsekas & Tsitsiklis, 1989; Olfati-Saber et al., 2007), and recent literature reviews (Cao et al., 2012; Kia et al., 2019). Most of these successful applications of cooperative control have been model-based.

[1]Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Bogotá, Colombia [2]Department of Electrical and Computer Engineering, Rice University, TX, USA.. Correspondence to: Miguel F. Arevalo-Castiblanco <miarevaloc@unal.edu.co>.

Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021.

Increasing theoretical understanding of the Reinforcement Learning (RL) methodologies has turned this framework into a practical option to develop data-based robust and efficient controllers (Recht, 2019). However, this is not an easy task as there are many challenges to consider when implementing these techniques in autonomous agents in real-life applications. One of the most critical challenges identified is the enormous variability of the results after multiple trials of the techniques on the same problem as presented in (Recht, 2019). The difference between the behaviors presented by the agents in simulation compared to the behaviors that occur in reality, called the *reality gap*, as presented in (Tan et al., 2018) and the cost (in time and energy) that it represents to test directly on the agent of interest compared to the cost required by the simulations (Koos et al., 2013).

Model reference adaptive control (MRAC) for leader-follower models have materialized as an important framework in adaptive control strategies for systems with uncertainty parameters (Nguyen et al., 2008; Nguyen, 2018) under online parameters adjustment.

In this paper, we propose a methodology for adaptive synchronization of heterogeneous agents using a learned policy. Leader-follower synchronization is achieved using a distributed MRAC (DMRAC) that improves the performance of a policy defined by an RL-trained algorithm. This policy is initially adjusted given the difference between a real system and a reference model. This development is integrated with a distributed reference-based framework for online-policy synchronization. A stability analysis using Lyapunov's theory is presented, guaranteeing the convergence of the proposed algorithm. Finally, simulation results for synchronizing a network of inverted pendulums network support our theoretical findings and allow its application in emulation systems.

The abstract is organized as follows. Section 2 introduces the optimal leader-follower synchronization problem. Section 3 shows the proposed distributed MRAC-RL and its stability analysis. Section 4 presents some simulation results and finally, in Section5 some conclusions are drawn.

2. Problem Formulation

In this section, we introduce a leader-follower synchronization with a reference model problem. Consider a set of N agents, represented as a continuous deterministic dynamic system of the form

$$\dot{x}_i = A_i x_i + b_i u_i, \qquad i \in [1, ..., N], \tag{1}$$

where $x_i \in \mathbb{R}^n$ are the agent states, $u_i \in \mathbb{R}^p$ is the control input, A_i is an unknown matrix associated to the agent states, b_i are known input vectors. The heterogeneity of the system is handled with $A_i \neq A_j$ and $b_i \neq b_j$. Moreover, define a reference model as

$$\dot{x}_m = A_m x_m + b_m u_m, \tag{2}$$

where $x_m \in \mathbb{R}^n$ is the reference state, and A_m and b_m are its states matrix and input vector, respectively. This system is associated with a cost functional for an optimal control problem defined in the supplementary file. Suppose we have *off-the-shelf* reinforcement learning strategies based on simulations on the reference model to generate a control policy $u_m(t) = \pi(x_m(t))$ to approximately solve optimization problem.

Remark 1. *Note that our goal is not to study the efficiency of RL controllers or to compare RL training methods. Instead, we seek to use a policy trained on a reference model, on a system with potentially different parameters.*

We define agents as a leader or leaders as the set of agents with access to the policy $\pi(\cdot)$, and the state and control action of the reference model, i.e., (u_m, x_m). Without loss of generality, we assume there is only one leader and denote it as agent 1.

Moreover, we assume all agents are connected over a network $\mathcal{G}(V, E)$, with $V = [1, \cdots, N]$ as the set of nodes or agents, and E as the set of edges, such that $(j, i) \in E$ if agent j is an in-neighbor of agent i. The adjacency matrix of the graph $\mathcal{G}$ is defined as $A = [a_{ij}]$ where $a_{ii} = 0$ and $a_{ij} = 1$ if $(j, i) \in E$, where $i \neq j$.

Assumption 1. *The communications graph $\mathcal{G}$ is unweighted, directed and, acyclic. The graph contains a directed spanning tree with the leader as the root node.*

A follower agent is defined as not having access to the policy $\pi(\cdot)$, nor the states or actions of the reference model. Follower agents can only observe the states and control actions of their in-neighbors on the network.

Assumption 1 guarantees that if synchronization on the states happens, all followers will eventually synchronize to the leader agent state.

Main objective: *Design a N control laws, $u_1 = g_1(x_m, u_m, x_1)$ and $u_i = g_i(x_i, \{u_j, x_j \mid (j, i) \in E\})$ for $i \in [2, \ldots, N]$. Such that, $e_1 = x_1 - x_m$ and $e_{ij} = x_i - x_j$ asymptotically go to zero. Figure 1 shows a representation of the communication structure and control of the main objective.*

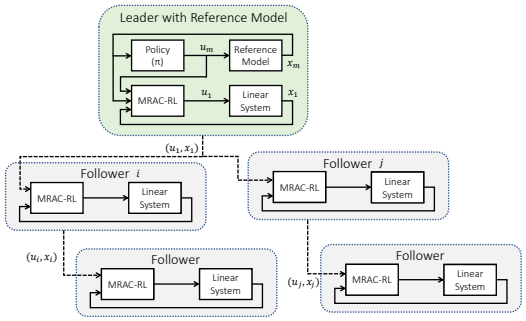

Figure 1. Block diagram DMRAC-RL with one leader and four followers.

3. Distributed Model Reference Adaptive Control with Reinforcement Learning

This section describes the proposed distributed multi-agent reinforcement learning model reference adaptive control, i.e., DMRAC-RL.

We consider a heterogeneous network of agents, where each agent is represented by dynamics (1). The reference model is described as in (2). In this specific case, each agent is modeled as a second-order system. We pick a set of diagonal matrices $D^i \in \mathbb{R}^{n \times n}$ with diagonal entries denoted in the supplementary file. The control law defined for the synchronization of the leader agent with respect to the reference model is defined as

$$u_1 = k_m x_m + k_r \xi, \tag{3}$$

where the adaptive gains $k_m \in \mathbb{R}^p \times \mathbb{R}^n$ is the constant associated with the reference states, $k_r \in \mathbb{R}^p \times \mathbb{R}^q$ is associated with the augmented reference signal $\xi := u_m - \frac{1}{b^1}(D^m \alpha_r^m)^\top e$, with D^m the diagonal matrix of the reference system. The adaptive laws are based on assumptions associated with matching conditions defined in the supplementary file and are denoted as

$$\dot{k}_m^\top = -\Gamma_m x_1^\top (x_1 - x_m) P_1 b_m^\top, \tag{4}$$

$$\dot{k}_r = -\gamma_r \xi (x_1 - x_m) P_1 b_m^\top. \tag{5}$$

where the adaptive gains $\Gamma_x \succ 0$ and $\gamma_u > 0$ are defined, and P can be obtained by a linear Lyapunov function. The proposition associated with this synchronization is found in the supplementary file along with the subsequent simplest case of a network of one leader and one follower.

In a general distributed case, the control law used for the synchronization of agents that do not have communication with the reference is

$$\bar{a}_i u_i = \sum_{j=1}^{N} a_{ij} k_{ij}^\top x_j + k_{mi}\Xi_i + \sum_{j=1}^{N} a_{ij} k_{rij}\xi_{ij},\ \forall i \in [2,\dots,N], \tag{6}$$

where $\Xi_i = \sum_{j=1}^{N} a_{ij}(x_i - x_j)$, $\bar{a}_i = \sum_{j=1}^{N} a_{ij}$ and $\xi_{ij} := u_j - \frac{1}{b^j}(D^j\alpha_r^j)^\top e_{ij}$. Using adaptive laws with $e_{ij} = (x_i - x_j)$ in the same way

$$\dot{k}_{ij}^\top = -\Gamma_{ij} x_j^\top \Xi_i P_i b_m^\top,$$
$$\dot{k}_{mi}^\top = -\Gamma_m \Xi_i^\top \Xi_i P_i b_m^\top, \tag{7}$$
$$\dot{k}_{rij} = -\gamma_r \xi_{ij} \Xi_i P_i b_m^\top. \tag{8}$$

Similarly, for follower agents $i \in [2,\dots,N]$, define P_i the matrix for which

$$P_i A_{Hi} + A_{Hi}^\top P_i = -Q_i, \qquad Q_i \succ 0, \tag{9}$$

where $A_{Hi} := A_m - h(D^i\alpha_r^i)^\top$.

Full proof can be found in Supplementary material Section 3, where an analysis of the selected Lyapunov function and its derivative is performed along the defined error dynamics e_{ij}. Moreover, we can state the main stability result of this synchronization problem in the following theorem.

Theorem 1. *Let Assumption 1 hold. The dynamics generated by the set of agents in (1), with control law (3) for the leader agent, and control law (6) for the followers, guarantee global uniform asymptotic synchronization, i.e.,* $\lim_{t\to\infty}\|e_{ij}(t)\| = 0$ *and* $\lim_{t\to\infty}\|e_1(t)\| = 0$*, with* $\|x_j(t)\| < M_{xj}\ \forall t \in [0,T]$ *for a constant* $M_{xj} > 0$*.*

Proof. Assuming that the reference signal $x_m(t)$ are bounded, from the described lemma it follows that the synchronization error e_{ij} and constants k_{mi}, k_{ij}, k_{rij} are uniformly bounded. The dynamics of the reference and the states are then also bounded, i.e., $x_j, \dot{x}_j, x_m, \dot{x}_m$ are bounded, and thus $x_i(t) = e_{ij} + x_j(t)$ is bounded, and at the same time it implies that $u_i(t)$ is bounded as well as $\dot{x}_i$ and $\dot{e}_{ij}$. To ensure uniform continuity of the Lyapunov function derivative, its second derivative is

$$\ddot{V} = -2\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} e_{ij}^\top Q_j e_{ij},$$

and is bounded because $V(t) \geq 0$ and $\dot{V}(t) \leq 0$. Thus, from Barbalat's Lemma we have that $\lim_{t\to\infty}\dot{V}(t) = 0$. Therefore, we can conclude that $\lim_{t\to\infty}\|e_{ij}(t)\| = 0$: the synchronization error tends to zero globally, asymptotically, and uniformly. □

Theorem 1 guarantees the synchronization of a network of heterogeneous agents to an RL-trained reference model.

4. Simulation Results

Consider the following linear and nonlinear model of an inverted pendulum

$$ml^2\ddot{\theta} = mgl\sin\theta - b\dot{\theta} + \tau, \tag{10}$$

where m is the pendulum mass, g is the gravitational constant, l is the length pendulum, τ is the force provided to the system. The goal is to maintain a non-zero set-point for the states $\theta, \dot{\theta}$. The linearized system around the equilibrium point is represented as

$$\dot{x}_i = \begin{bmatrix} 0 & 1 \\ \frac{g}{l_i} & \frac{b}{m_i l_i^2} \end{bmatrix} x_i + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_i. \tag{11}$$

We use an off-the-shelf *Deep Deterministic Policy Gradient Agent* pre-trained policy from MATLAB®. This policy was trained to swing up and balance an inverted pendulum. Parameter and training process details can be found in (Inc., 2021). Figure 2 shows a trained pendulum simulation using the mentioned reinforcement learning toolbox from MATLAB different variations of the parameters. Remarkably, the percentage value indicates the absolute deviation from the nominal parameters in the reference model. The pre-trained policy stabilizes the pendulum around the equilibrium point for the reference model. However, when the linear system parameters are different from those used in the training phase, the system might not converge to the equilibrium point.

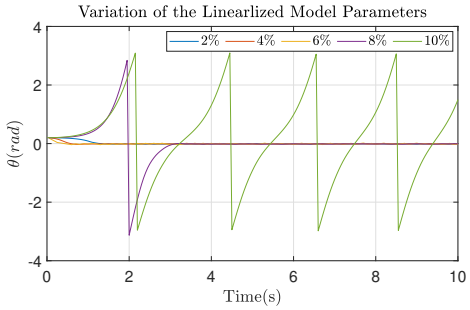

Figure 2. Nonlinear pendulum response with reinforcement learning strategy.

In the Supplementary material Section 4 we show experiments, to test the efficiency of the MRAC-RL technique when the parameters of a follower system are modified, as well as its initial conditions based on Figure 2. The test is made for cases where the reference model has linear and non-linear dynamics. Similarly, to validate the distributed MRAC-RL strategy, we initially used a 6-node communication graph, where its response to a reference model with linear and non-linear characteristics is evidenced. Here, we

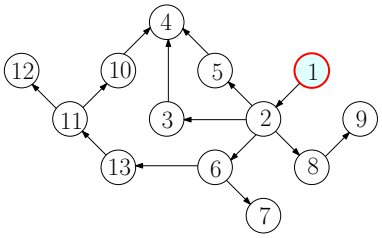

Figure 3. Distributed communication network, a red border denotes the lead agent.

show the response along the communication network of Figure 3 with 12 nodes in which Figure 4 show the trajectories generated by the network of pendulums with different initial conditions. These figures do not show a legend by space but the trajectory of all the pendulums of the network is plotted. The initial state is π radians, i.e., facing downwards.

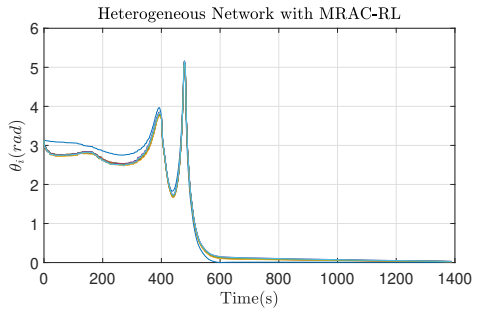

Figure 4. Nonlinear reference with heterogeneous pendulum with larger network and distributed MRAC-RL with initial conditions modification.

5. Conclusions

We proposed a distributed MRAC framework for the adaptive synchronization of networks of leader-followers heterogeneous agents using a control policy defined by an RL-trained algorithm. The proposed DMRAC-RL uses an internal loop that adjusts the policy for agents directly and complements an external loop in an augmented input to solve the distributed control problem. A stability analysis using Lyapunov's theory has been presented to guarantee the convergence of the proposed algorithm. Numerical experiments, including linear and nonlinear reference models for a network of pendulums, have been carried out, and leader-follower synchronization has been achieved. Future work should study the inclusion of disturbances in the model and extensions to systems nonlinear dynamics both for the leader and followers.

References

Bertsekas, D. P. and Tsitsiklis, J. N. Parallel and distributed computation: numerical methods, 1989.

Cao, Y., Yu, W., Ren, W., and Chen, G. An overview of recent progress in the study of distributed multi-agent coordination. *IEEE Transactions on Industrial informatics*, 9(1):427–438, 2012.

Inc., M. Train ddpg agent to swing up and balance pendulum. Mathworks Reference page, 2021. Accessed: 2021-03-24.

Kia, S. S., Van Scoy, B., Cortes, J., Freeman, R. A., Lynch, K. M., and Martinez, S. Tutorial on dynamic average consensus: The problem, its applications, and the algorithms. *IEEE Control Systems Magazine*, 39(3):40–72, 2019. doi: 10.1109/MCS.2019.2900783.

Koos, S., Mouret, J., and Doncieux, S. The transferability approach: Crossing the reality gap in evolutionary robotics. *IEEE Transactions on Evolutionary Computation*, 17(1):122–145, 2013. doi: 10.1109/TEVC.2012.2185849.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Lewis, F. L., Zhang, H., Hengster-Movric, K., and Das, A. *Cooperative control of multi-agent systems: optimal and adaptive design approaches*. Springer Science & Business Media, 2013.

Nguyen, N. *Model-Reference Adaptive Control. A Primer*. Springer, March 2018.

Nguyen, N., Krishnakumar, K., and Boskovic, J. An optimal control modification to model-reference adaptive control for fast adaptation. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, pp. 7283, 2008.

Olfati-Saber, R., Fax, J. A., and Murray, R. M. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007. doi: 10.1109/JPROC.2006.887293.

Recht, B. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2(1):253–279, 2019. doi: 10.1146/annurev-control-053018-023825.

Tan, J., Zhang, T., Coumans, E., Iscen, A., Bai, Y., Hafner, D., Bohez, S., and Vanhoucke, V. Sim-to-real: Learning agile locomotion for quadruped robots. *ArXiv preprints.*, 2018.

Wang, Y., Garcia, E., Zhou, Z., Kingston, D., and Casbeer, D. *Cooperative control of multi-agent systems*. Wiley Online Library, 2017.