2021 Seventh Indian Control Conference (ICC)
IIT Mumbai, India. December 20-22, 2021

A Law of Iterated Logarithm for Multi-Agent Reinforcement Learning

Gugan Thoppe
Computer Science and Automation
Indian Institute of Science
Bengaluru, Karnataka 560012, India
Email: gthoppe@iisc.ac.in

Bhumesh Kumar
Electrical and Computer Engineering
University of Wisconsin at Madison
Madison, WI 53706, USA
Email: bkumar@wisc.edu

***Abstract*—In Multi-Agent Reinforcement Learning (MARL), multiple agents interact with a common environment, as also with each other, for solving a shared problem in sequential decision-making. In this work, we derive a novel law of iterated logarithm for a family of distributed nonlinear stochastic approximation schemes that is useful in MARL. In particular, our result describes the convergence rate on almost every sample path where the algorithm converges. This result is the first of its kind in the distributed setup and provides deeper insights than the existing ones, which only discuss convergence rates in the expected or the CLT sense. Importantly, our result holds under significantly weaker assumptions: neither the gossip matrix needs to be doubly stochastic nor the stepsizes square summable.**

I. Introduction

An archetypical setup of MARL has a directed graph $\mathcal{G}$ with m distributed nodes and a matrix $W \equiv (W_{ij}) \in [0,1]^{m\times m}$ whose ij-th entry denotes the strength of the edge $j \to i$ in $\mathcal{G}$. The update rule at agent i is given by

$$x_{n+1}(i) = \sum_{j\in\mathcal{N}_i} W_{ij}x_n(j) + \alpha_n[h_i(x_n) + M_{n+1}(i)], \quad (1)$$

where $x_n \in \mathbb{R}^{m\times d}$ is the joint estimate at time n, its j-th row, i.e., $x_n(j)$ denotes[1] the estimate obtained at agent j, $\mathcal{N}_i$ represents the set of in-neighbors of node i in $\mathcal{G}$, α_n is the stepsize, $h_i : \mathbb{R}^{m\times n} \to \mathbb{R}^d$ is the driving function at agent i, and $M_{n+1}(i) \in \mathbb{R}^d$ is the noise in the evaluation of $h_i(x_n)$.

The joint update rule of all the agents is

$$x_{n+1} = Wx_n + \alpha_n[h(x_n) + M_{n+1}], \quad (2)$$

where M_{n+1} is the $m\times d$ matrix whose i-th row is $M_{n+1}(i)$, and h is the function that maps $x \in \mathbb{R}^{m\times d}$ to the $m \times d$ matrix whose i-th row is $h_i(x)$. Let $\mathcal{E}(x_*) := \{x_n \to x_*\}$.

II. Assumptions and Main Result

We make the following four technical assumptions, i.e., $\mathcal{A}_1, \ldots, \mathcal{A}_4$, in relation to the DSA scheme given in (2).

- $\mathcal{A}_1$) **Property of the Gossip Matrix**: *W is an irreducible aperiodic row stochastic matrix.*
- $\mathcal{A}_2$) **Nature of h near x_***: *There exists a neighbourhood $\mathcal{U}$ of x_* such that, for $x \in \mathcal{U}$,*

$$h(x) = -\mathbf{1}'\pi(x-x_*)A + \mathbf{1}'\pi f_1(x) + (\mathbb{I}-\mathbf{1}'\pi)(B+f_2(x)),$$

where A and f_1 have certain regularity properties.

- $\mathcal{A}_3$) **Stepsize Behaviour**: *There exists some decreasing positive function α defined on $[0,\infty)$ such that the stepsize $\alpha_n = \alpha(n)$. Further, α is either of Type 1 or Type γ.*
- $\mathcal{A}_4$) **Noise Attributes**: *With $\mathcal{F}_n = \sigma(x_0, M_1, \ldots, M_n)$ and M_{n+1} is martingale difference sequence with certain regularity properties.*

These regularity conditions are detailed in [1] and generalize the standard assumptions [2], [3], [4]. Our main result can now be stated as follows. This generalizes Theorem 1 from [2].

Theorem II.1 (Main Result: Law of Iterated Logarithm). *Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_4$ hold and $\gamma > 2/b$. Then, there exists some deterministic constant $C \geq 0$ such that*

$$\limsup [\alpha_n \log t_{n+1}]^{-1/2} \|x_n - x_*\| \leq C \quad \text{a.s. on } \mathcal{E}(x_*).$$

Remark II.2. *Our result shows that, a.s. on $\mathcal{E}(x_*)$, $\|x_n - x_*\|$ is $\mathcal{O}(\sqrt{n^{-1}\log\log n})$ in the Type 1 case, and $\mathcal{O}(\sqrt{n^{-\gamma}\log n})$ in the Type γ case.*

III. Proof of the Main Result: A Sketch

The key steps in the proof of Theorem II.1 follow. The complete details can be found in [1]. With $Q = \mathbb{I} - \mathbf{1}'\pi$, observe that $\mathbf{1}'\pi x_* = x_*$ and, hence, $x_n - x_* = \mathbf{1}'\pi(x_n - x_*) + Qx_n$. The first and second terms are referred as the agreement and disagreement components. This decomposition differs from the standard approaches [5], [6], [7], wherein $x_n - x_*$ is split into $(\mathbf{1}'\mathbf{1}/m)(x_n - x_*)$ and $(\mathbb{I} - (\mathbf{1}'\mathbf{1}/m))x_n$.

Lemma III.1. *(Agreement Error) Almost surely on $\mathcal{E}(x_*)$,*

$$\limsup_{n\to\infty} \frac{\|\mathbf{1}'\pi(x_n - x_*)\|}{\sqrt{\alpha_n \log t_{n+1}}} \leq C. \quad (3)$$

Lemma III.2. *(Disagreement Error) Almost surely on $\mathcal{E}(x_*)$, for any $\delta > 0$*

$$\limsup_{n\to\infty} \frac{\|Qx_n\|}{\alpha_n(\log n)^{1+\delta}} \leq C. \quad (4)$$

IV. Future Directions

While our DSA framework is fairly general, the limitation on matrix A, dynamic communication protocols and two-timescale DSA schemes hold potential for future research.

[1] By default, all our vectors are row vectors. We use $'$ for transpose.

2021 Seventh Indian Control Conference (ICC) | 978-1-6654-0978-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC54714.2021.9702912

ACKNOWLEDGMENT

We would like to thank Prof. Vivek Borkar for suggesting this exciting problem. We would also like to thank the anonymous reviewers for providing helpful and constructive feedback on the paper. Research of Gugan Thoppe is supported by IISc's start up grants SG/MHRD-19-0054 and SR/MHRD-19-0040.

REFERENCES

[1] G. Thoppe and B. Kumar, "A law of iterated logarithm for multi-agent reinforcement learning," *arXiv preprint arXiv:2110.15092*, 2021.

[2] M. Pelletier, "On the almost sure asymptotic behaviour of stochastic algorithms," *Stochastic processes and their applications*, vol. 78, no. 2, pp. 217–244, 1998.

[3] A. Mokkadem and M. Pelletier, "Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms," *The Annals of Applied Probability*, vol. 16, no. 3, pp. 1671–1702, 2006.

[4] V. S. Borkar, *Stochastic approximation: a dynamical systems viewpoint*. Springer, 2009, vol. 48.

[5] T. Doan, S. Maguluri, and J. Romberg, "Finite-time analysis of distributed td (0) with linear function approximation on multi-agent reinforcement learning," in *International Conference on Machine Learning*, 2019, pp. 1626–1635.

[6] T. T. Doan, S. T. Maguluri, and J. Romberg, "Finite-time performance of distributed temporal-difference learning with linear function approximation," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 298–320, 2021.

[7] G. Morral, P. Bianchi, and G. Fort, "Success and failure of adaptation-diffusion algorithms with decaying step size in multiagent networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2798–2813, 2017.