Two-Timescale Q-Learning with Function Approximation in Zero-Sum Stochastic Games

ZAIWEI CHEN, California Institute of Technology, USA
KAIQING ZHANG, University of Maryland, College Park, USA
ERIC MAZUMDAR, California Institute of Technology, USA
ASUMAN OZDAGLAR, Massachusetts Institute of Technology, USA
ADAM WIERMAN, California Institute of Technology, USA

We consider two-player zero-sum stochastic games and propose a two-timescale variant of Q-learning with function approximation that is payoff-based, convergent, rational, and symmetric between the two players. In two-timescale Q-learning, the fast-timescale iterates are updated in spirit to the stochastic gradient descent for minimizing a Bellman error variant and the slow-timescale iterates (which we use to compute the policies) are updated by taking a convex combination between its previous iterate and the latest fast-timescale iterate. In the special case of linear function approximation, we present, to the best of our knowledge, the first last-iterate finite-sample bound for payoff-based independent learning dynamics of these types.

To establish the results, we analyze our proposed algorithm using a two-timescale stochastic approximation framework, and derive the finite-sample bound through a Lyapunov-based approach. The key technical novelty lies in the construction of a valid Lyapunov function to capture the evolution of the slow-timescale iterates. Specifically, through a change of variable, we show that the update equation of the slow-timescale iterates resembles the classical smoothed best-response dynamics, where the regularized Nash gap serves as a valid Lyapunov function. This insight enables us to construct a valid Lyapunov function via a generalized variant of the Moreau envelope of the regularized Nash gap. The construction of our Lyapunov function might be of independent interest in studying the dynamics of general stochastic approximation algorithms.

The full paper is publicly available at https://arxiv.org/abs/2312.04905.

CCS Concepts: • **Computing methodologies** → **Multi-agent reinforcement learning**; • **Theory of computation** → **Stochastic approximation**.

Additional Key Words and Phrases: Zero-Sum Stochastic Games, Q-Learning, Last-Iterate Convergence, Lyapunov Function

ACM Reference Format:
Zaiwei Chen, Kaiqing Zhang, Eric Mazumdar, Asuman Ozdaglar, and Adam Wierman. 2024. Two-Timescale Q-Learning with Function Approximation in Zero-Sum Stochastic Games. In *The 25th ACM Conference on Economics and Computation (EC '24), July 8–11, 2024, New Haven, CT, USA.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3670865.3673491

Authors' Contact Information: Zaiwei Chen, zchen458@caltech.edu, California Institute of Technology, Pasadena, CA, USA; Kaiqing Zhang, kaiqing@umd.edu, University of Maryland, College Park, MD, USA; Eric Mazumdar, mazumdar@caltech.edu, California Institute of Technology, Pasadena, CA, USA; Asuman Ozdaglar, asuman@mit.edu, Massachusetts Institute of Technology, Cambridge, MA, USA; Adam Wierman, adamw@caltech.edu, California Institute of Technology, Pasadena, CA, USA.

EC '24, July 8–11, 2024, New Haven, CT, USA

ACM ISBN 979-8-4007-0704-9/24/07
https://doi.org/10.1145/3670865.3673491