Learning in Multi-Memory Games Triggers Complex Dynamics Diverging from Nash Equilibrium

Yuma Fujimoto
Research Center for Integrative Evolutionary Science,
SOKENDAI (The Graduate University for Advanced Studies).
Universal Biology Institute (UBI), the University of Tokyo.
CyberAgent, Inc.
fujimoto_yuma@soken.ac.jp

Kaito Ariu
CyberAgent, Inc. / KTH
kaito_ariu@cyberagent.co.jp

Kenshi Abe
CyberAgent, Inc.
abe_kenshi@cyberagent.co.jp

Abstract

Repeated games consider a situation where multiple agents are motivated by their independent rewards throughout learning. In general, the dynamics of their learning become complex. Especially when their rewards compete with each other like zero-sum games, the dynamics often do not converge to their optimum, i.e., Nash equilibrium. To tackle such complexity, many studies have understood various learning algorithms as dynamical systems and discovered qualitative insights among the algorithms. However, such studies have yet to handle multi-memory games (where agents can memorize actions they played in the past and choose their actions based on their memories), even though memorization plays a pivotal role in artificial intelligence and interpersonal relationship. This study extends two major learning algorithms in games, i.e., replicator dynamics and gradient ascent, into multi-memory games. Then, we prove their dynamics are identical. Furthermore, theoretically and experimentally, we clarify that the learning dynamics diverge from the Nash equilibrium in multi-memory zero-sum games and reach heteroclinic cycles (sojourn longer around the boundary of the strategy space), providing a fundamental advance in learning in games.

1 Introduction

Repeated game models that multiple agents aim to optimize their objective functions based on a normal-form game [1]. It is known that in this game, the set of optimal strategies for all the agents always exists as Nash equilibria [2]. Various algorithms with which each agent achieves its optimal strategy have been proposed, such as Cross learning [3], replicator dynamics [4, 5], gradient ascent [6, 7, 8, 9], Q-learning [10, 11, 12], and so on. In zero-sum games where two agents have conflicts in their benefits, however, the above learning algorithms cannot converge to their equilibrium [13, 14]. Indeed, the dynamics of learning draw a loop around the equilibrium point, even though the shape of the trajectory differs more or less depending on the algorithm. Thus, solving the dynamics around the Nash equilibrium is a touchstone for discussing whether the learning works well.

Currently, several studies attempt to understand trajectories of multi-agent learning by integrating various cross-disciplinary algorithms [15, 16, 17, 18]. For example, if we take an infinitesimal step size of learning, Cross learning draws the same trajectory as a replicator dynamics. The replicator dynamics can be interpreted as the weighted version of infinitesimal gradient ascent. Furthermore, Q-learning differs only in the extra term of exploration with the replicator dynamics. Another study has shown a relationship between the replicator dynamics and Q-learning by introducing a generalized regularizer which pulls the strategy back to the probabilistic simplex at the shortest distance [13]. Like these studies, it is important to understand the trajectory of multi-agent learning theoretically.