Neural Population Learning beyond Symmetric Zero-sum Games

Siqi Liu
Google DeepMind and
University College London
London, United Kingdom
liusiqi@google.com

Luke Marris
Google DeepMind and
University College London
London, United Kingdom
marris@google.com

Marc Lanctot
Google DeepMind
Montréal, Canada
lanctot@google.com

Georgios Piliouras
Google DeepMind
London, United Kingdom
gpil@google.com

Joel Z. Leibo
Google DeepMind
London, United Kingdom
jzl@google.com

Nicolas Heess
Google DeepMind
London, United Kingdom
heess@google.com

ABSTRACT

We study computationally efficient methods for finding equilibria in n-player general-sum games, specifically ones that afford complex visuomotor skills. We show how existing methods would struggle in this setting, either computationally or in theory. We then introduce NeuPL-JPSRO, a neural population learning algorithm that benefits from transfer learning of skills and converges to a Coarse Correlated Equilibrium (CCE) of the game. We show empirical convergence in a suite of OpenSpiel games, validated rigorously by exact game solvers. We then deploy NeuPL-JPSRO to complex domains, where our approach enables adaptive coordination in a MuJoCo control domain and skill transfer in capture-the-flag. Our work shows that equilibrium convergent population learning can be implemented at scale and in generality, paving the way towards solving real-world games between heterogeneous players with mixed motives.

KEYWORDS

Game Theory; Deep Learning; Multiagent Reinforcement Learning; Coarse Correlated Equilibrium

ACM Reference Format:
Siqi Liu, Luke Marris, Marc Lanctot, Georgios Piliouras, Joel Z. Leibo, and Nicolas Heess. 2024. Neural Population Learning beyond Symmetric Zero-sum Games. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

1 INTRODUCTION

Purely competitive, symmetric zero-sum games have proven to be popular testbeds for AI research since its early days [5, 6, 28, 31, 34, 35, 38, 39]. Principled algorithms have been developed in this setting, with convergence guarantees to a Nash Equilibrium (NE, [26]) where players can be expected to win (or draw) against *any* opponent. One family of equilibrium convergent methods follows from Fictitious Play (FP, [4]) and Double Oracle (DO, [23]). By learning a set of strategies each best-responding to a mixture over their predecessors, FP and DO converge to an NE even in cyclic games (e.g. *rock-paper-scissors*) where self-play would have made no progress. Policy-Space Response Oracle (PSRO, [14]) extended similar guarantees to extensive-form (EF, [12]) games by constructing a normal-form (NF) metagame whose actions correspond to playing a deep reinforcement learning (RL) policy for an entire episode. Variations of this idea have led to competitive agents in games as complex as StarCraft [39], albeit at significant costs training and evaluating hundreds of independent deep RL agents.

While significant advances have been made in finding Nash Equilibria in symmetric zero-sum games, real-world interactions are often n-player general-sum — between heterogeneous actors with mixed motives. Progress here has been more limited for a few reasons. Computationally, finding exact NE is intractable beyond two-player zero-sum games (i.e. PPAD-complete [9]). More importantly, NE describe an impoverished view of general-sum interactions as it forbids correlated action choices between players. This limitation is subtle but critical: consider a road junction, an NE can only suggest uncorrelated action choices for each driver when much improved outcomes could have been achieved by coordinating drivers with a trusted third-party (e.g. a traffic light). Similar general-sum interactions occur frequently in our society. Fair, mutually beneficial social norms often enable coordination and improve outcomes for all parties.

This observation motivated (Coarse) Correlated Equilibria ((C)CE, [1, 24]), an equilibrium solution concept that allows for coordinated actions between players, mediated by a correlation device that *rational*, *self-interested* players would find beneficial to follow (go on "green", wait on "red"). (C)CE generalise NE, as they naturally reduce to NE if players are in pure competition and have no way to usefully coordinate. Unlike NE, (C)CE are computationally tractable too, as they can be formulated as a linear program (LP) even in the n-player general-sum setting. Algorithms that offer convergence to (C)CE have also received increased interest in recent years. Following similar iterative best-response arguments as PSRO, [21] proposed Joint PSRO (JPSRO), a population learning algorithm with convergence guarantees to a NF (C)CE in n-player general-sum EF games. Nevertheless, evidence of convergence to (C)CE has been limited to a few research games that can be solved analytically — the costs of representing, training and evaluating a population of independent RL agents *for each player* quickly become intractable, especially in games that demand complex skills.

How can we bring game-theoretic algorithms such as (J)PSRO to real-world games in full generality and at scale? The central

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).