Discrete-time Heavy-Anchor Reinforcement Learning in Multiagent Finite Games

Supervisor: Prof. Lacra Pavel

Student: Rupin Khadwal

Student Number: 1006163232

Table of Contents

1. Introduction

This investigation aims to explore the potential development of a reinforcement-learning (RL) scheme that converges to the true Nash Equilibrium (NE) in multiagent finite games. Building upon the work of Bolin Gao and Lacra Pavel in their 2021 paper, which utilized passivity based approach but achieved convergence to a perturbed NE[1]. Their work on Q-learning research introduces a different RL scheme, P-RL [2]. Furthermore, research on heavy anchor dynamics and Q-Learning with side-information proved convergence to the true NE can be attained in monotone games [3] [4]. While Gao's and Pavel's model demonstrated convergence in games characterized by the monotonicity property of negative payoff vectors, the proposed approach seeks to find a solution in a discrete-time system with side-information as most systems are implemented in such time.

2. Literature Review:

Introduction

This investigative journey ambitiously embarks on the mission of developing an advanced reinforcement-learning (RL) scheme aimed at achieving the authentic Nash Equilibrium (NE) in the complex landscape of multiagent finite games. Building upon the pioneering work of Gao and Pavel, this novel approach incorporates heavy anchor dynamics and Q-Learning with Side-information to assess the feasibility of attaining convergence to the true NE within a discrete-time system. The quest for NE-seeking in non-cooperative games propels our purpose [3][4]. RL, with its adaptable nature and minimal informational requirements compared to traditional methods like fictitious play and gradient play, emerges as a potential solution. The overarching goal expands beyond mere convergence, delving into the unexplored realms of RL's applicability in scenarios marked by incomplete information. The purpose is dual-fold: extending Gao's model and proposing innovative RL approaches, with a distinct emphasis on convergence within discrete-time systems, aligning with the practical implementation of most systems.

Exploring the basic concepts of Game Theory and Nash Eqiulibrum seeking, will deliver a base understanding of important notions to help move into advanced and specific concepts. Thereafter, the narrative unfurls into the intricate web of four pivotal components: Passivity-

Based Controls for NE Seeking, P-RL & Q-Learning in Continuous Time, Heavy Anchor Dynamics, and Q-Learning with Side Information. The organizational architecture is more than a structural framework; it is a deliberate narrative that unfolds coherently, revealing the interconnectedness of theories, methodologies, and innovations. The comprehensive roadmap extends from foundational concepts to detailed analyses of each critical component, providing readers with a holistic view of our pursuit for true NE convergence in the dynamic landscape of multiagent finite games. A few important definitions are also mentioned.

Definitions

Nash Equilibrium: The point in a non-cooperative game wherein players have no incentive to unilaterally deviate from their strategy. [5]

Reinforcement Learning: An area of machine learning that focuses on rewarding/punishing desired/undesired behaviors to "teach" the agent an optimal policy. [6]

Continuous-Time System: A continuous-time system is a dynamical system where the input, output, and state variables are defined at every instant of time in a continuous manner. The system evolves smoothly and continuously over time. [7]

Discrete-Time System: A discrete-time system is a dynamical system in which the input, output, and state variables are defined at distinct, separate time instances. The system evolves in discrete steps or intervals, with time progressing in a quantized manner. [8]

The Motivation to use RL in NE-Seeking

Nash equilibrium-seeking is a challenging problem without a general solution for non-cooperative games. Reinforcement learning (RL) in multiagent finite games, especially those with incomplete information, has gained interest due to its weak informational requirements.

While RL shows potential for solving games that other methods cannot, challenges remain. Prior research mostly focused on convergence results in potential games and two-player zero-sum games. Gao et al.'s research addressed this gap, proposing that passivity-based control theory could make existing RL schemes converge in N-player monotone and hypomonotone games.[1]

Game Theory

Game Theory, a branch of applied mathematics, provides a robust framework for analyzing situations characterized by interdependent decisions among multiple parties, known as players. The essence of game theory lies in unraveling the strategic intricacies that unfold when players formulate decisions, considering the potential moves of others. Originating from the collaborative efforts of John von Neumann, a Hungarian-born American mathematician, and Oskar Morgenstern, a German-born American economist, game theory was initially conceptualized to address challenges in economics. Their seminal work, "The Theory of Games and Economic Behavior" [9], argued for the inadequacy of traditional physical sciences mathematics in capturing the strategic dynamics inherent in economic interactions. Instead, they proposed game theory as a novel mathematical approach suited to the nuanced decision-making processes involved in economic activities. [10]

Game theory encompasses scenarios where players may have similar, opposed, or mixed interests, leading to a diverse array of potential outcomes. The strategic considerations in decision-making, as opposed to pure chance, set game theory apart from classical probability theory. Its applications extend far beyond traditional parlour games, permeating various fields such as politics, business, pricing strategies, voting dynamics, jury selection, and even ecological studies of animal and plant behaviors. The theory aids in predicting and understanding the formation of political coalitions, determining optimal pricing strategies in competitive markets, assessing the power dynamics of voters or voter blocs, and optimizing decisions related to manufacturing plant locations. [10]

The versatility of game theory is evident in its application to challenges ranging from legal disputes about voting systems to the optimal placement of manufacturing plants. It has been instrumental in shedding light on the dynamics of strategic interactions in various contexts, offering valuable insights into decision-making processes influenced by complex interdependencies. [10]

NE-Seeking

At the core of game theory lies the concept of Nash Equilibrium (NE), a pivotal outcome in noncooperative games for multiple players. Coined after the eminent American mathematician John Nash, the NE represents a state where no player can enhance their expected outcome by unilaterally altering their strategy. This foundational idea serves as a cornerstone in game theory, especially in N-player noncooperative games, earning Nash the 1994 Nobel Prize in Economics for his ground-breaking contributions [11] [12] [13].

Crucial to understanding NE-seeking is the classification of games as noncooperative, meaning players lack mechanisms for binding agreements. A classic example is the prisoner's dilemma, where two accused individuals face the dilemma of confessing or remaining silent without any enforceable agreement. The absence of external enforcement renders the game noncooperative, emphasizing the strategic nature of decisions where betrayal incurs no penalty. [14]

Understanding when and where this state occurs, along with predicting player payoffs at that point, is crucial in competitive game theory. While algorithms like fictitious play and gradient play have been devised to find the NE, they are limited by informational requirements. RL algorithms, explored in detail, have gained prominence due to their applicability in games with limited information, where traditional algorithms fall short.

Passivity-Based Controls for NE Seeking

The scrutiny of Reinforcement Learning (RL) within the realm of passivity-based control theory unfolds a tapestry of substantial contributions and breakthroughs, elucidating intricate facets of convergence in multiagent finite games. The cardinal contributions articulated in the study can be expounded upon to unveil a more nuanced understanding of the advancements in RL.

The utilization of passivity-based control theory serves as a cornerstone in establishing the convergence of an existing RL scheme. The study illuminates that this convergence is not limited to potential games but encompasses a broader spectrum, specifically extending to N-player monotone and hypomonotone games. The significance of this extension, lies in the broadened applicability of RL in diverse game scenarios, transcending previous constraints and providing a more encompassing solution for convergence challenges. [1]

Expanding the horizon of passivity-based control, the research introduces the concept of higher-order learning dynamics, ushering in a paradigm shift in the design of RL extensions. By delving into the integration of higher-order dynamics through auxiliary states, the study underscores the potential advantages in fostering convergence for expansive classes of games, effectively surmounting the limitations posed by conventional first-order schemes. [1]

As depicted in Figure 1, there are still limitations in convergence with passivity-based approaches that still need to be addressed, specifically with monotone games. The Anti-coordination is a benchmark strategic game which cannot be resolved via traditional passivity-based approaches and needs further research to be solved.

The study further extends its purview to discrete-time reinforcement learning, particularly focusing on a scheme with noisy updates grounded in the stochastic-approximation method. The revelations encapsulated in Theorem 3 underscore the adaptability and convergence capabilities of passivity-based control in discrete-time settings. This extension enhances the practicality and versatility of RL algorithms, catering to real-world scenarios where continuous-time implementations may be impractical, thereby broadening the scope of RL applications. [1]

In essence, this deep dive into passivity-based control theory within the ambit of RL not only broadens the scope of convergence but also pioneers a framework for the design of higher-order learning dynamics. These developments offer a sophisticated and nuanced comprehension of RL's potential in navigating complex decision spaces involving multiple agents. The robust theoretical foundations, fortified by meticulous proofs and numerical results, contribute substantively to the scientific dialogue, paving the way for more resilient and adaptable RL applications across diverse scenarios and challenges.[1]

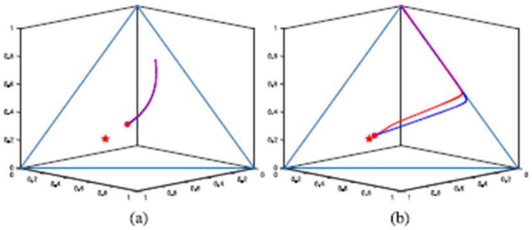


Figure 1. The convergence of the 123 Anticoordination game using the EXP-D-RL approach. (a) *epsilon* = 1. (b) *epsilon* = 0.1. Taken from [1]

P-RL & Q-Learning in Continuous Time

Understanding the importance of P-RL and Q-Learning methods in continuous time is vital. This passage discusses two interconnected systems (P) represented by equations (1) and (2). These systems are visualized in Figures 2(a) and 2(b), with the key difference lying in the configuration of the forward path. Figure 2(a) involves a bank of integrators, while Figure 2(b) incorporates a bank of low-pass filters. The properties of R in these figures are examined, revealing that R in Figure 2(a) exhibits Extended Input Passivity (EIP), while R in Figure 2(b) demonstrates Output Strictly Extended Input Passivity (OSEIP).

Further analysis is carried out to leverage OSEIP passivity for the asymptotic stability of the Q-learning closed-loop system (equation 2) shown in Figure 2(b). The discussion emphasizes the generality of the results, asserting stability for any epsilon greater than zero in any N-player monotone game, where the monotonicity of the negative payoff game mapping (-U) plays a crucial role. Additionally, the passage explores the limitations of the P-RL system in Figure 2(a), where only mere stability is achievable when -U is monotone, highlighting the importance of passivity techniques to extend convergence results to a broader class of hypomonotone games and design higher-order Q-learning dynamics. The goal is to strike a balance between the passivity characteristics on the feedback and feedforward paths, ensuring convergence to an approximated Nash Equilibrium for any epsilon greater than zero.

$$P\colon \begin{cases} \dot{z} = U_i(x) - z, \;\; z(0) \in \mathbb{R}^n \\ x = \sigma_\epsilon(z) \end{cases}$$

Equation 1. Q-Learning Feeback System

$$P\colon \begin{cases} \dot{z} = U_i(x), \;\; z(0) \in \mathbb{R}^n \\ x = \sigma_\epsilon(z) \end{cases}$$

Equation 2. P-RL Feeback System

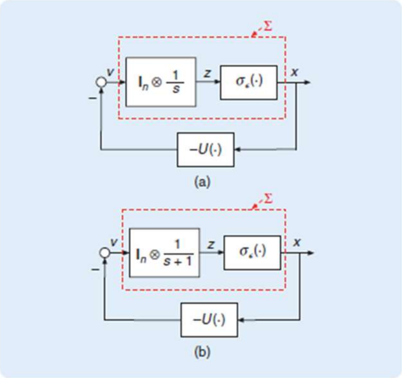
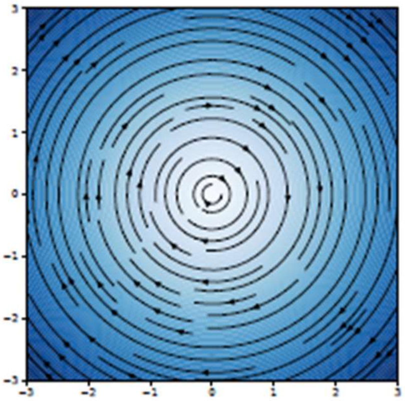


Figure 2. a) Payoff-based reinforcement learning (P-RL) and (b) Q-learning, represented as a feedback interconnected system (R,*U*), where *U* on the feedback path is the payoff game mapping. On the forward path, R is the composition between (a) a bank of integrators (P-RL) or (b) a bank of low-pass filters (Q-learning) and the soft-max mapping. Taken from [2]

Heavy Anchor Dynamics

In this comprehensive investigation, the focus is on advancing the understanding and applicability of distributed Nash equilibrium seeking within the intriguing domains of monotone and hypomonotone games. The paper unveils an innovative solution named "Heavy Anchor," which stands out as a passivity-based modification of the conventional gradient-play dynamics. The primary objective of Heavy Anchor is to overcome the strict monotonicity constraints of the pseudo-gradient, which is imperative for gradient-play dynamics. The paper rigorously proves that Heavy Anchor achieves not only a relaxation of strict monotonicity but also ensures exact asymptotic convergence in merely monotone regimes, a pivotal contribution that extends the reach of convergence results beyond conventional boundaries. [3]

The study takes a bold step forward by extending the applicability of Heavy Anchor to scenarios where players possess only partial information about their opponents' decisions. In this setting, each player maintains a local decision variable and an auxiliary state estimate, fostering a decentralized learning approach. The modification of Heavy Anchor through distributed Laplacian feedback becomes instrumental in leveraging equilibrium-independent passivity properties to attain convergence to a Nash equilibrium, particularly in hypomonotone regimes. These findings mark a significant leap in the literature, Figure 3 and Figure 4 show the integration of Heavy Anchor in a monotone gradient field. [3]



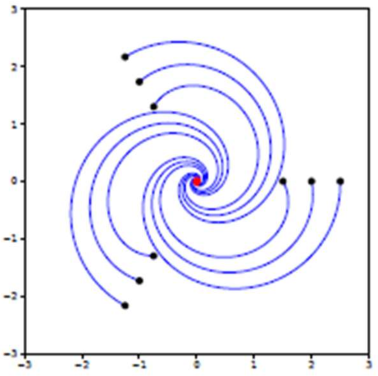
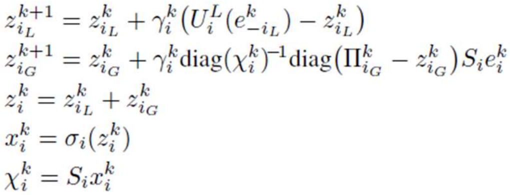


Figure 3. Gradient Vector Field. Taken from [3] Figure 4. Decision trajectories under Heavy Anchor. Taken from [3]

The contributions of the paper are multifaceted. First and foremost, it introduces and rigorously analyzes the Heavy Anchor algorithm, showcasing its prowess in ensuring exact convergence to a Nash equilibrium for positive parameter values in the full-decision information setting. Moreover, the extension of Heavy Anchor to hypomonotone games, under specified conditions, underscores its adaptability to more complex scenarios. In the partial-decision information setting, the paper achieves a notable milestone by proving convergence for monotone extended pseudo-gradient or hypomonotone and inverse Lipschitz pseudo-gradient, filling a significant gap in the existing literature. The exploration of Heavy Anchor's relationship to optimization dynamics, its similarity to approaches used in chaotic systems stabilization, and its connections to second-order dynamics in the optimization realm further enrich the scientific discourse. Overall, Heavy Anchor emerges as a versatile and powerful methodology, offering novel insights and solutions to long-standing challenges in distributed Nash equilibrium seeking.[3]

Q-Learning with Side Information

The paper introduces a discrete-time Nash equilibrium-seeking reinforcement learning scheme designed to exploit side information, ultimately achieving convergence in a specific class of finite games characterized by negative monotonicity properties in their utility. Notably, the research emphasizes the limited literature on reinforcement learning that effectively utilizes side information despite its potential practical applications. In response to this gap, the study explores various Q-Learning techniques, including FLQL, IQL, and QLSI 1-3, all derived from central equations (Figure 5) but featuring distinct simplifications based on initial conditions and assumptions that converge at different rates as depicted in Figure 6. [4]



$$
\begin{aligned}
z_{i_L}^{k+1} &= z_{i_L}^{k} + \gamma_i^k \big(U_i^L(e_{-i_L}^k) - z_{i_L}^k\big) \\
z_{iG}^{k+1} &= z_{iG}^{k} + \gamma_i^k \operatorname{diag}(\chi_i^k)^{-1} \operatorname{diag}\big(\Pi_{iG}^k - z_{iG}^k\big) S_i e_i^k \\
z_i^k &= z_{i_L}^k + z_{iG}^k \\
x_i^k &= \sigma_i(z_i^k) \\
\chi_i^k &= S_i x_i^k
\end{aligned}
$$

Figure 5. Q-learning with Side Information (QLSI) updating functions. Taken from [4]

The contributions of the paper are multifaceted, aligning with the growing field of multi-agent reinforcement learning. The algorithm's design adheres to the minimal information assumptions of classical multi-agent reinforcement learning algorithms, ensuring compatibility with existing frameworks. However, it introduces a more general formulation, allowing players to harness additional side information when available. Two distinct forms of side information are considered: a local/global split in utility functions and the exploitation of payoff measurements for unplayed actions. The local/global split model, akin to previous work, provides full information for the local component while minimizing information on the global utility part. In contrast to correlated equilibrium-seeking algorithms, this algorithm specifically focuses on seeking Nash equilibrium. The second form of side information involves knowledge of the payoff a player would have obtained had it chosen a different action, adding a layer of strategic depth to the learning process. Through numerical simulations of representative games, the paper showcases that exploiting more side information leads to a faster convergence rate to a Nash equilibrium, highlighting the practical significance of incorporating side information in reinforcement learning for finite games. [4]

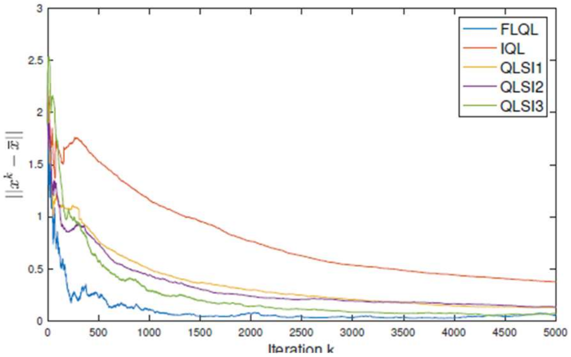


Figure 6. Distance from the Nash equilibrium as a function of the iteration obtained with different algorithms for the standard RPS game. Taken from [4]

Conclusion

This literature review embarks on advancing reinforcement-learning (RL) schemes for achieving the Nash Equilibrium (NE) in multiagent finite games. Building on Gao and Pavel's Q-learning approach, our research introduces the modified P-RL scheme, incorporating heavy anchor dynamics and Q-Learning with side-information for convergence in discrete-time systems. The purpose extends beyond convergence, exploring RL's applicability in games with incomplete information. The existing gap involves challenges in NE-seeking, with Gao et al. proposing passivity-based control theory for N-player monotone and hypomonotone games.

3. Current Work

The current research efforts encompass a comprehensive exploration of NE-seeking techniques, specifically focusing on the base EXP-D RL model derived from Lacra's paper. The completed solutions include the discrete-time implementation of Q-Learning, P-RL, and P-RL with Heavy Anchor Dynamics, emphasizing numerical approximations to ensure accuracy. Extensive visualization of simulation data, covering convergence patterns and numerical errors, has been executed to gain a holistic understanding of the model's performance. This work extends to the mathematical derivation of discrete-time transfer functions, with the P-RL system already successfully formulated. Ongoing efforts involve determining the discrete-time transfer functions for Q-Learning and Heavy Anchor Dynamics. The entire research process is meticulously documented on Google Colab (ipynb) for potential collaboration and ease of access, underscoring the commitment to transparency and collaborative engagement.

The foundational work completed thus far includes recreating Lacra's base EXP-D RL model in discrete-time, fostering a deep understanding of NE-seeking techniques. Close collaboration with Lacra has facilitated a nuanced comprehension of the intricate details of the proposed methods. The visual representation of simulation outcomes, covering convergence trajectories and numerical discrepancies, provides a robust basis for analysis. Additionally, discrete-time transfer functions have been successfully derived for the P-RL system, contributing to the mathematical underpinning of the research. The ongoing investigations into the discrete-time transfer functions for Q-Learning and Heavy Anchor Dynamics signify a continued dedication to a thorough exploration of NE-seeking methodologies. The comprehensive documentation on Google Colab (ipynb) not only ensures transparency but also facilitates potential collaborative endeavors and streamlines accessibility for interested parties.

4. Future work

In the upcoming phases of the research, the focus will be on achieving precise discrete-time transfer functions for both Q-Learning and Heavy Anchor Dynamic feedback loop. This involves a meticulous mathematical analysis to derive exact transfer functions, moving beyond numerical approximations. The subsequent step involves the integration of these precise discrete-time functions into Python for implementation, marking a shift towards a more accurate and computationally efficient modeling approach. Furthermore, the exploration of incorporating Q-Learning with Side Information into the existing system will be initiated. This endeavor aims to identify optimal placements within the system for the integration of side information, shedding light on its impact and potential enhancements in achieving Nash Equilibrium.

Anticipated outcomes include the attainment of nearly exact Nash Equilibrium for the P-RL system and the P-RL system with Heavy Anchor Dynamics, specifically in the Anti-coordination game. This aligns with the overarching goal of refining and validating the models to mirror real-world scenarios more accurately. With the introduction of side-information, improved convergence is expected, offering a deeper understanding of the system dynamics influenced by additional information. The prospect of achieving relatively faster convergence further highlights the potential benefits of incorporating side-information, showcasing its positive impact on the learning and decision-making processes within the context of reinforcement learning.

References

1. B. Gao and L. Pavel, "On Passivity, Reinforcement Learning, and Higher Order Learning in Multiagent Finite Games," IEEE Transactions on Automatic Control, vol. 66, pp. 121-136, 2021.

2. L. Pavel, "Dissipativity Theory in Game Theory," IEEE Control Systems, pp. 150-164, 2022.

3. D. Gadjov and L. Pavel, "On the exact convergence to Nash equilibrium in hypomonotone regimes under full and partial-information," Proc. 59th IEEE Conf. Decision Control, pp. 2297-2302, 2020.

4. M. Sylvestre and L. Pavel, "Q-Learning with Side Information in Multi-Agent Finite Games," 2019 IEEE 58th Conference on Decision and Control (CDC), Nice, France, 2019, pp. 5032-5037, doi: 10.1109/CDC40024.2019.9029788.

5. O. Chatain, "Cooperative and Non-Cooperative Game Theory," in The Palgrave Encyclopedia of Strategic Management, 2014.

6. L. P. Kaelbling, M. L. Littman and A. W. Moore, "Reinforcement Learning: A Survey," Journal of Artifcial Intelligence Research, vol. 4, pp. 237-285, 1996.

7. K. Ogata, "System Dynamics," Pearson, 2010.

8. A. V. Oppenheim and R. W. Schafer, "Discrete-Time Signal Processing," Prentice Hall, 1999.

9. J. von Neumann and O. Morgenstern, "The Theory of Games and Economic Behavior," Princeton University Press, 1944.

10. M. Davis and S. Brams, “Game theory,” Encyclopædia Britannica, https://www.britannica.com/science/game-theory

11. J. Nash, "Non-cooperative Games," Annals of Mathematics, vol. 54, no. 2, pp. 286-295, 1951.

12. J. Nash, "The Imbedding Problem for Riemannian Manifolds," Annals of Mathematics, vol. 63, no. 1, pp. 20-63, 1956.

13. J. Nash, "The imbedding problem for differential manifolds," Annals of Mathematics, vol. 63, no. 1, pp. 20-63, 1956.

14. S. Eldridge, “Nash equilibrium,” Encyclopædia Britannica, https://www.britannica.com/science/Nash-equilibrium.

List of Figures & Tables