Reinforcement Learning in Multi-Objective Multi-Agent Systems

Doctoral Consortium

Willem Röpke
Artificial Intelligence Lab
Vrije Universiteit Brussel
Belgium
willem.ropke@vub.be

ABSTRACT

For effective decision-making in the real world, artificial agents need to take both the multi-agent as well as multi-objective nature of their environments into account. These environments are formalised as multi-objective games and introduce numerous challenges compared to their single-objective counterpart. For my main contributions so far, I have established a theoretical guarantee that a bidirectional link always exists that maps a finite multi-objective game to an equivalent single-objective game with an infinite number of actions. Additionally, I presented an extensive study of Nash equilibria in multi-objective games, culminating in existence guarantees under certain assumptions. From a reinforcement learning perspective, I explored how communication and commitment can help agents to learn adequate policies in these challenging environments. In this paper, I summarise my ongoing research and discuss several promising directions for future work.

KEYWORDS

Multi-objective; Game theory; Reinforcement learning

ACM Reference Format:
Willem Röpke. 2023. Reinforcement Learning in Multi-Objective Multi-Agent Systems: Doctoral Consortium. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

To be effective in real-world settings, artificial agents must be able to navigate complex decision-making scenarios. In many settings of interest, this is complicated by the presence of multiple agents with whom cooperation and competition is possible [21]. To study decision-making in these types of systems, game theorists define solution concepts which are joint strategies that are stable in some sense. A well-known solution concept is the Nash equilibrium, in which agents have no incentive to unilaterally deviate from the joint strategy and independently play their strategy without any means of communication or correlation [6].

Multi-agent reinforcement learning (MARL), on the other hand, focuses specifically on the acting aspect of decision-making. Agents interact with their environment and each other with the goal of learning adequate policies, possibly belonging to an equilibrium. MARL and game theory are closely linked, with advances in one area often benefiting the other [5, 18].

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

An additional challenge in decision-making is the presence of multiple conflicting objectives, particularly when agents are designed to represent the interests of humans. This is further amplified in multi-agent settings, where each agent may have distinct preferences or even different objectives. For example, a scientist writing a paper may have conflicting goals of making the paper concise yet thorough and novel yet centred in a relevant area. Determining the optimal trade-off between these objectives can be difficult, and the preferred solution may vary depending on the individual. The field of multi-objective decision-making targets such problems and provides methods to deal with this complexity [11].

Multi-objective games combine the multi-objective and multi-agent aspects of decision-making, returning a vector-valued payoff instead of a scalar [1]. A popular model is the multi-objective normal-form game (MONFG), which generalizes the classic normal-form game. To deal with the vectorial payoffs, it is common to accept a utility-based approach which assumes the existence of a utility function for each agent [3, 10]. Note, however, that the utility function need not be public knowledge or given a priori. As such, it is often impossible or undesirable to reduce the multi-objective to a single-objective game [8].

It is known that different optimisation criteria naturally arise in the utility-based approach. Agents may optimise on the basis of expected utility, as is commonly assumed in game theory, which leads to the expected scalarised returns (ESR) criterion. On the other hand, research in multi-objective reinforcement learning usually considers agents that optimise for the scalarised expected returns (SER) criterion, i.e. settings where the utility is derived from the expected payoff [3]. It is known that the choice of optimisation criteria influences which strategies are optimal [19] and even when certain equilibria can be guaranteed to exist [9].

Despite progress in related fields, multi-objective games are largely unexplored, with a lack of results regarding the existence of equilibria and efficient computational approaches for learning or computing optimal strategies. Given their relevance in modelling real-world settings, continued research on these aspects is vital.

2 RELATING MULTI-OBJECTIVE TO SINGLE-OBJECTIVE GAMES

In games where agents aim to optimise their expected utility, i.e. the ESR criterion, it is possible to reduce the multi-objective game to a single-objective game when utility functions are known. Intuitively, this is because the utility function can be applied to the payoff vectors, resulting in an equivalent single-objective game. When optimising for SER, however, it has been shown that this is not possible when agents have nonlinear utility functions. Consider for

example the payoffs $(2, 0)$ and $(0, 2)$ and the product utility function $u(p_1, p_2) = p_1 \cdot p_2$. For a uniform mixture over the two payoffs, the utility of the expected payoff is equal to one but is different than the expected utility of the payoffs which is equal to zero. Moreover, for nonlinear utility functions, it is not guaranteed that a Nash equilibrium exists, which stands in stark contrast to results from single-objective games [9]. This is particularly relevant as humans often have nonlinear utility functions [7, 16]. To better understand the dynamics of these games and connect them to other established concepts in game theory, it is therefore crucial to further investigate the properties of MONFGs.

In recent work, we introduce a novel equivalence notion, called pure strategy equivalence, that relates MONFGs to single-objective games with continuous action spaces, referred to as continuous games [15]. We showed that a bidirectional link can be constructed, which maps a game from one class to the other while preserving underlying dynamics such as Nash equilibria. As a consequence, it becomes possible to translate theoretical results from continuous games to MONFGs. Additionally, as MONFGs can be represented in a succinct matrix format with a finite number of actions, it opens the possibility of designing algorithms that leverage this structure to tackle continuous games. We demonstrate this by designing a fictitious play algorithm for multi-objective games and using this to learn Nash equilibria in different continuous games.

3 NASH EQUILIBRIA IN MULTI-OBJECTIVE GAMES

The existence and computation of Nash equilibria is a central question in game theoretic research. In multi-objective games, this question has been explored in settings with unknown utility functions, leading to the introduction of Pareto Nash equilibria [4, 17]. When utility functions are known, previous work has shown that MONFGs under ESR can be reduced to single-objective games [9]. For games under SER, however, comparatively little is known.

Recently, we contributed an extensive study of Nash equilibria in MONFGs [13]. In our work, we derive conditions such that existence is guaranteed under SER and computation becomes feasible. Specifically, we show that Nash equilibria can be guaranteed to exist when assuming only quasiconcave utility functions. Such utility functions are a generalisation of concave functions and imply that agents prefer an average payoff for all objectives over extremes [2]. On the other hand, we showed that no such guarantee is possible for strict convex utility functions, thereby also precluding existence guarantees under convexity or quasiconvexity.

As computing Nash equilibria under ESR is feasible by first performing the reduction to a single-objective game, it is useful to find conditions such that an equilibrium under ESR is also one under SER. Unfortunately, we demonstrate that no general relation between ESR and SER exists. However, when restricting our attention to pure strategy Nash equilibria, i.e. equilibria where strategies are deterministic, we find that equilibria are shared when agents have quasiconvex utility functions. Additionally, we show that this extends to settings where a subset of agents optimises for ESR while others optimise for SER. Finally, we contributed an efficient algorithm that combines these results to compute all pure strategy Nash equilibria.

4 LEARNING WITH COMMUNICATION

In challenging environments, it may be infeasible to compute equilibrium strategies a priori. Additionally, even when computation is feasible, there may be multiple equilibrium strategies with no obvious choice for which equilibrium to play. In MARL, these challenges are addressed by enabling agents to learn which strategies to play in response to the other agents in the environment.

In multi-objective games, independent actor-critic learning has been proposed with a modified objective function to directly optimise the utility from the expected returns [22]. We extend this approach by introducing communication protocols that encourage agents to learn adequate policies through iterative play of the MONFG [14]. In each iteration, one agent is designated as the leader and the other as a follower. The leader is then required to commit to some strategy after which the follower is allowed to condition their response on the commitment, analogous to the mechanism of a Stackelberg game [20].

We introduce variations for collaborative and self-interested agents, where each variation prescribes what type of commitment the leader can make and how the follower may react. First, we present two cooperative protocols where the leader either communicates their next action or current policy and the follower performs a policy update based on this information. Notably, agents are forced to lead and follow with the same policy. While stable policies may not exist, we find that these protocols punish agents that deviate too much from a suitable middle ground and thus encourage cooperation.

Next, we presented a self-interested variant where agents are allowed to learn distinct best-response policies for each commitment and lead with a different policy than when following. We observed that agents cycled through two policies, one when leading and one when following. Interestingly, these learned policies may constitute cyclic equilibria. We performed a follow-up study where we demonstrated that cyclic equilibria can be rational in MONFGs [12].

5 FUTURE WORK

The overarching goal of my research is to develop a comprehensive understanding of multi-objective games and design computational techniques that can effectively compute or learn equilibria. With this goal in mind, there are a number of interesting areas I aim to explore for future work.

First, I believe that the novel bidirectional link between MONFGs and continuous games presents exciting opportunities for further research. I aim to study the design of general algorithms for MONFGs and investigate the possibility of applying these methods to the solution of complex continuous games or finding approximate solutions through convenient approximate MONFG representations. Additionally, Stackelberg games with convex strategy sets have been studied in the literature [20]. I aim to explore how the link can provide novel insights and contributions to this area of research.

Furthermore, I aim to study techniques that enable decision support in multi-objective games where agents represent human interests. As it is often difficult for humans to exactly specify their utility function, it will be necessary to develop interactive solutions that can learn both the preferences of the user and optimal play in the game.

ACKNOWLEDGMENTS

Willem Röpke is supported by the Research Foundation – Flanders (FWO), grant number 1197622N.

REFERENCES

[1] David Blackwell. 1954. An Analog of the Minimax Theorem for Vector Payoffs. *Pacific J. Math.* 6, 1 (1954), 1–8. https://doi.org/10.2140/pjm.1956.6.1

[2] John W. Fowler, Esma S. Gel, Murat M. Köksalan, Pekka Korhonen, Jon L. Marquis, and Jyrki Wallenius. 2010. Interactive Evolutionary Multi-Objective Optimization for Quasi-Concave Preference Functions. *European Journal of Operational Research* 206, 2 (2010), 417–425. https://doi.org/10.1016/j.ejor.2010.02.027

[3] Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. 2022. A Practical Guide to Multi-Objective Reinforcement Learning and Planning. *Autonomous Agents and Multi-Agent Systems* 36, 1 (April 2022), 26. https://doi.org/10.1007/s10458-022-09552-y

[4] Anisse Ismaili. 2018. On Existence, Mixtures, Computation and Efficiency in Multi-Objective Games. In *PRIMA 2018: Principles and Practice of Multi-Agent Systems*, Tim Miller, Nir Oren, Yuko Sakurai, Itsuki Noda, Bastin Tony Roy Savarimuthu, and Tran Cao Son (Eds.). Springer International Publishing, Cham, 210–225.

[5] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Perolat, David Silver, and Thore Graepel. 2017. A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning. In *Advances in Neural Information Processing Systems*, I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett (Eds.), Vol. 30. Curran Associates, Inc., 4190–4203.

[6] John Nash. 1951. Non-Cooperative Games. *The Annals of Mathematics* 54, 2 (1951), 286–286. https://doi.org/10.2307/1969529

[7] Joost M. E. Pennings and Ale Smidts. 2003. The Shape of Utility Functions and Organizational Behavior. *Management Science* 49, 9 (2003), 1251–1263.

[8] Roxana Rădulescu, Patrick Mannion, Diederik M. Roijers, and Ann Nowé. 2020. Multi-Objective Multi-Agent Decision Making: A Utility-Based Analysis and Survey. *Autonomous Agents and Multi-Agent Systems* 34, 1 (April 2020), 10–10. https://doi.org/10.1007/s10458-019-09433-x

[9] Roxana Rădulescu, Patrick Mannion, Yijie Zhang, Diederik M. Roijers, and Ann Nowé. 2020. A Utility-Based Analysis of Equilibria in Multi-Objective Normal-Form Games. *The Knowledge Engineering Review* 35 (2020), e32–e32. https://doi.org/10.1017/S0269888920000351

[10] Diederik M. Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A Survey of Multi-Objective Sequential Decision-Making. *Journal of Artificial Intelligence Research* 48 (2013), 67–113. https://doi.org/10.1613/jair.3987

[11] Diederik M. Roijers and Shimon Whiteson. 2017. Multi-Objective Decision Making. In *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 34. Morgan and Claypool, 129–129. https://doi.org/10.2200/S00765ED1V01Y201704AIM034

[12] Willem Röpke, Roxana Radulescu, Ann Nowé, and Diederik M. Roijers. 2022. Commitment and Cyclic Strategies in Multi-Objective Games. In *Proceedings of the Adaptive and Learning Agents Workshop (ALA 2022)*, Francisco Cruz, Conor F. Hayes, Felipe Leno da Silva, and Fernando P. Santos (Eds.). Online, \url{https://ala2022.github.io/}, 9.

[13] Willem Röpke, Diederik M. Roijers, Ann Nowé, and Roxana Rădulescu. 2022. On Nash Equilibria in Normal-Form Games with Vectorial Payoffs. *Autonomous Agents and Multi-Agent Systems* 36, 2 (Oct. 2022), 53. https://doi.org/10.1007/s10458-022-09582-6

[14] Willem Röpke, Diederik M. Roijers, Ann Nowé, and Roxana Rădulescu. 2022. Preference Communication in Multi-Objective Normal-Form Games. *Neural Computing and Applications* (July 2022). https://doi.org/10.1007/s00521-022-07533-6

[15] Willem Röpke, Carla Groenland, Roxana Rădulescu, Ann Nowé, and Diederik M. Roijers. 2023. Bridging the Gap Between Single and Multi Objective Games. https://doi.org/10.48550/ARXIV.2301.05755

[16] Michael Scholz, Verena Dorner, Markus Franz, and Oliver Hinz. 2015. Measuring Consumers' Willingness to Pay with Utility-Based Recommendation Systems. *Decision Support Systems* 72 (2015), 60–71. https://doi.org/10.1016/j.dss.2015.02.006

[17] Kiran K. Somasundaram and John S. Baras. 2009. Achieving Symmetric Pareto Nash Equilibria Using Biased Replicator Dynamics. In *Proceedings of the IEEE Conference on Decision and Control*. IEEE, Shanghai, China, 7000–7005. https://doi.org/10.1109/CDC.2009.5400799

[18] Karl Tuyls and Gerhard Weiss. 2012. Multiagent Learning: Basics, Challenges, and Prospects. In *AI Magazine*, Vol. 33. 41–52. https://doi.org/10.1609/aimag.v33i3.2426

[19] Peter Vamplew, Cameron Foale, and Richard Dazeley. 2022. The Impact of Environmental Stochasticity on Value-Based Multiobjective Reinforcement Learning. *Neural Computing and Applications* 34, 3 (Feb. 2022), 1783–1799. https://doi.org/10.1007/s00521-021-05859-1

[20] Bernhard von Stengel and Shmuel Zamir. 2010. Leadership Games with Convex Strategy Sets. *Games and Economic Behavior* 69, 2 (2010), 446–457. https://doi.org/10.1016/j.geb.2009.11.008

[21] Michael Wooldridge. 2009. *An Introduction to MultiAgent Systems* (second ed.). John Wiley & Sons.

[22] Yijie Zhang, Roxana Rădulescu, Patrick Mannion, Diederik M. Roijers, and Ann Nowé. 2020. Opponent Modelling for Reinforcement Learning in Multi-Objective Normal Form Games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Auckland, New Zealand, 2080–2082–2080–2082.