Machine Learning Project: part 1

Ignace Bleukx Quinten Bruynseraede

March 21, 2020

1 Literature Review

In this first assignment we try to combine basic principles from game theory with the work concerning multi-agent reinforcement learning. Most literature included in this literature review will therefore more of less fall into one of these categories. First we give an overview of the relevant literature, along with their contributions.

Article	Contribution
Multi-agent systems: Algorithmic, Game-Theoretic,and Logical Foundations, Shoham and Leyton-Brown [8]	This paper provides a thorough explaination of the different aspect of game theory, including different types of equilibria. These concepts are of importance to us since we will investigate whether or not our learning algorithms converge to one of these equilibria. Furthermore, the paper provides a detailed description of different types of games, such as cooperative games and non-cooperative games, as well as the notion of games in normal form.
Multi-agent learning dynamics, Bloembergen [2]	This thesis on multi agent learning dynamics provides essential information about different game theoretic aspects. Mainly section 2.3 on evolutionary game theory and chapter 3 on learning dynamics are relevant. In this last chapter, the replicator dynamics of many matrix games are investigated and explained very clearly. In this chapter we find an example of the learning pattern we would like to observe with our application of different learning algorithms.
OpenSpiel: A Framework for Reinforcement Learning in Games, Lanctot et al. [6]	The paper provides the documentation of the OpenSpiel framework. All aspects of the library are explained, from installation to implemented algorithms and games. Many design choices of the framework are clarified which helps to understand the philosophy behind the framework. In the paper, the game theoretic aspects are briefly touched upon, as well as important concepts of the implemented learning algorithms. This paper is of very much importance to us as we will use (and potentially extend) the OpenSpiel framework for this assignment.
Reinforcement learning produces dominant strategies for the Iterated Prisoner's Dilemma, Harper et al. [5]	This document contains a detailed description of the prisoners dilemma, which we will examine. Some examples of parameters settings for the training algorithms are given, which will help to produce meaningful results when training the learning algorithms of choice.
The replicator equation on graphs, Ohtsuki and Nowak [7]	This paper provides an insight on the visualization of the replicator dynamics using phase plots, as well as some examples relevant to our research. These examples include the prisoners dilemma and biased rock-paper-scissors.
Analyzing Reinforcement Learning algorithms using Evolutionary Game Theory, Bloembergen [1]	This thesis provides a rich source of information on the reinforcement learning branch of evolutionary game theory. Many algorithms are examined, some of which are available in OpenSpiel. The paper also contains the exact parameter settings used to reproduce its results. These parameters can be used by our agents to achieve similar results.
Evolutionary Dynamics of Multi-Agent Learning:A Survey, Bloembergen et al. [3]	Like other papers, this document provides a basic knowledge of game theory, as well as reinforcement learning. For our research, mainly the part about lenient FAQ-learning as a way to increase the robustness of Q-learning, is important. FAQ-learning is able to recover from bad exploration in the start of the run, while normal Q-learning is sometimes not.

Extended Replicator Dynamics as a Key to Reinforcement Learning in Multi-agent Systems, Tuyls et al. [10]	To model stochastic policies, populations of players are used. These populations can be described using evolutionary concepts, such as selection and mutation. This paper explains the transition from regular to evolutionary game theory. We received insight on the dynamics of a population through the central notion of replicator dynamics. These selection mechanisms can be extended with mutation, based on the Boltzmann mechanism. To overcome converge to suboptimal equilibria, lenience towards mistakes is introduced in this paper.

2 Independent learning

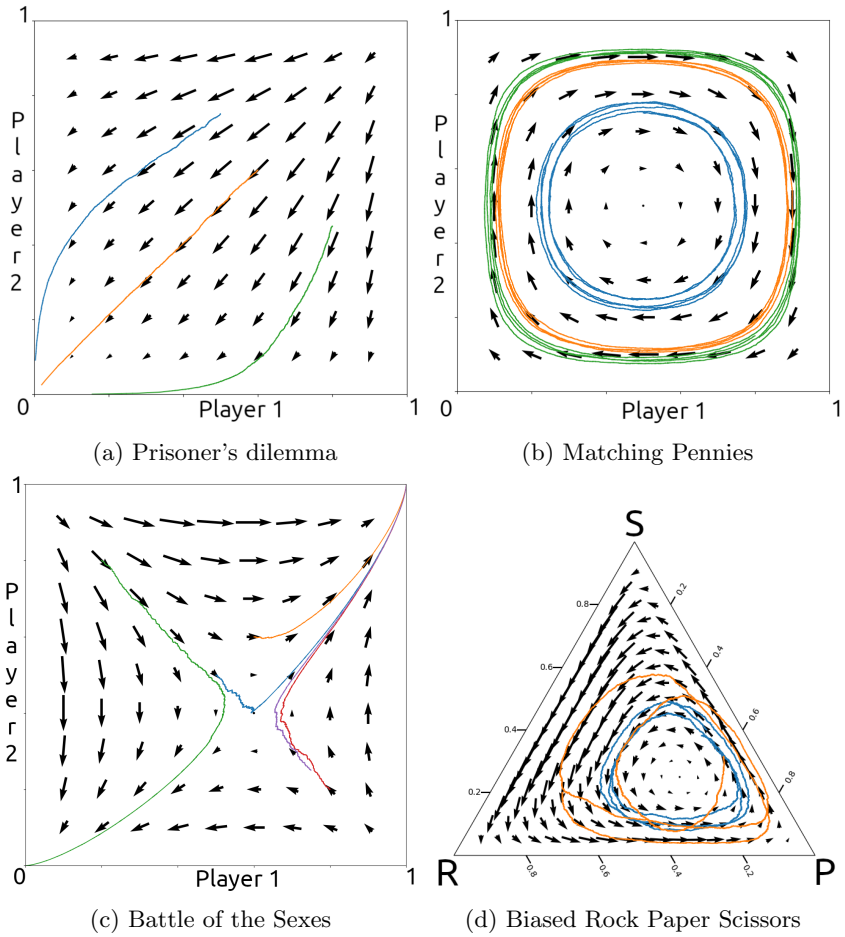

Figure 1: Learning trajectories in benchmark games

For this section we implemented our own reinforcement learning agent which follows the *Cross Learning* algorithm. Börgers & Saring were the first to formally show the relationship between multi-agent learning and evolutionary game theory by proving that Cross Learning converges to the replicator dynamics when the agent's step size goes towards 0. For this reason, we chose implement Cross Learning: we can expect its learning trajectories to follow our directional plot more or less exactly. In Figure 1, we show how our algorithm learns four games: the prisoner's dilemma, the matching pennies game, the battle of the sexes, and biased Rock-Paper-Scissors. Also included in Figure 1 are the directional phase plots for replicator dynamics. Each plot contains learning trajectories from multiple initial policies.

Cross Learning converges to $(0, 0)$ for the Prisoner's Dilemma, the situation in which both prisoners defect. This is a Nash equilibrium, as both players will lower their reward by cooperating while the other player still defects. It is also Pareto optimal, because when one player suddenly decides to cooperate, he himself receives a lower reward. It is therefore not possible to improve a strategy without lowering the reward for another player.

In the Matching Pennies game, the algorithm converges to $(\frac{1}{2}, \frac{1}{2})$, which is a Nash equilibrium: if a player decides to change his strategy whilst the other one keeps playing heads or tails randomly, he will be at a disadvantage. As Matching Pennies is a zero-sum game, all strategies are Pareto-optimal: there is a set amount of reward to distribute, so improving a player's reward lowers the reward for another player.

A known result from Game Theory is that the Battle of the Sexes games has three equilibria, one of which is mixed. The strategies $(1, 1)$ and $(0, 0)$ are pure Nash equilibria, while $(\frac{3}{5}, \frac{2}{5})$ is a mixed Nash equilibrium. $(1, 1)$ and $(0, 0)$ are Pareto optimal, because changes in policies have the following effects: when one player changes his strategy, both players go to a different activity and receive no reward. When both players change their strategy, one player doesn't go to his preferred activity anymore. The mixed strategy is also Pareto optimal: going to your preferred solution more often lowers the reward of your partner, etc. This mixed strategy is unstable.

For the Biased Rock Paper Scissors game, we observe one equilibrium in the directional phase plots, at approximately $(\frac{1}{4}, \frac{2}{4}, \frac{1}{4})$. Regular Rock Paper Scissors has its Nash equilibrium at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, but changing the payoff matrix moves this equilibrium towards a certain action. As Rock Paper Scissors is essentially Matching Pennies with three actions, it is a zero-sum game. Therefore, any strategy profile is Pareto optimal, using the same argument as before.

All experiments were executed for 30 000 iterations, a learning rate of 0.001. To demonstrate the correlation between Cross Learning and replicator dynamics, we did not average out results of multiple experiments. Instead, we kept the learning rate as low as possible.

3 Dynamics of learning

3.1 Lenient Boltzmann Q-Learning dynamics

Based on Bloembergen et al. [3], we implemented Lenient Boltzmann Q-Learning dynamics. This extension introduces two parameters. The first parameter, κ, is the degree of leniency: the number of rewards that are examined in each step before updating the policy. We expect to see improved robustness when introducing leniency. Intuitively, this is done by increasing the area of attraction for optimal equilibria. Secondly, τ introduces entropy into the population: a learner will favour exploration over exploitation when τ increases. We expect fixed points of the dynamics to stray away from Nash equilibria as τ is increased. However, introducing a small amount of entropy may prevent attraction from unwanted equilibria early on.

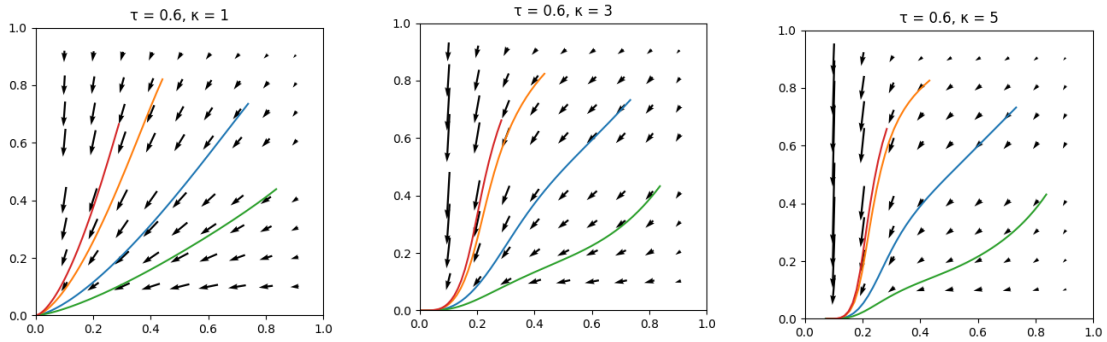

Figure 2: LFAQ for the Prisoner's dilemma

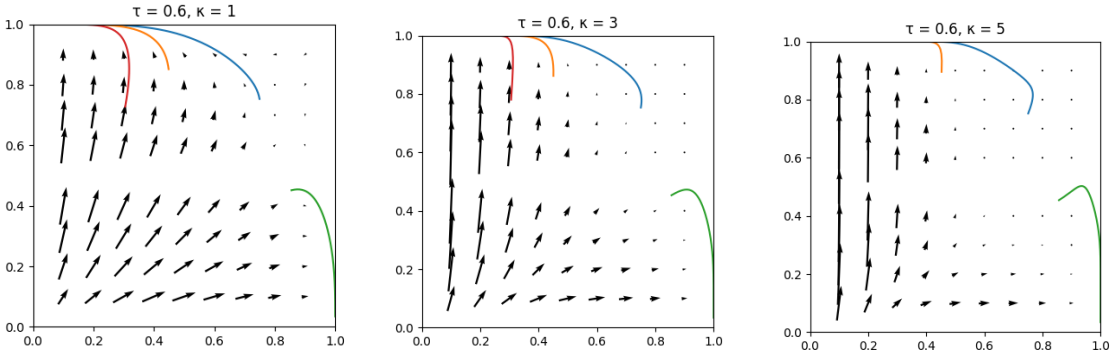

Figure 3: LFAQ for the Battle of the Sexes

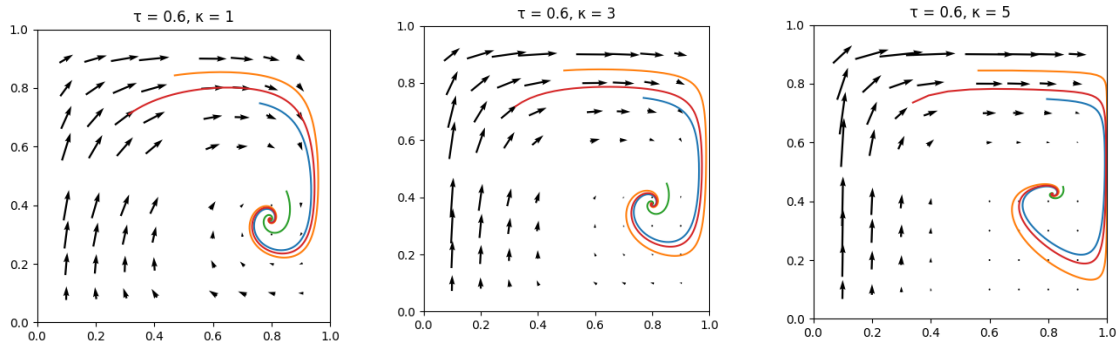

Figure 4: LFAQ for the Matching Pennies game

Figures 2 through 4 show the influence of the parameter κ on the dynamics. For games such as the Prisoner's Dilemma with one equilibrium (in this case located at 0,0), the difference isn't very interesting. However, in Figure 3, we clearly see how there is no longer attraction to the suboptimal equilibrium at $(\frac{3}{5}, \frac{2}{5})$.

The influence of τ is very clear, as shown in Figure 5. As τ increases, exploration of new solutions is greatly preferred. Therefore, an agent will not further capitalize on improvements in his strategy. As a result, policies seem to converge to random guessing. A small of amount of entropy, however, may increase robustness of a learner.

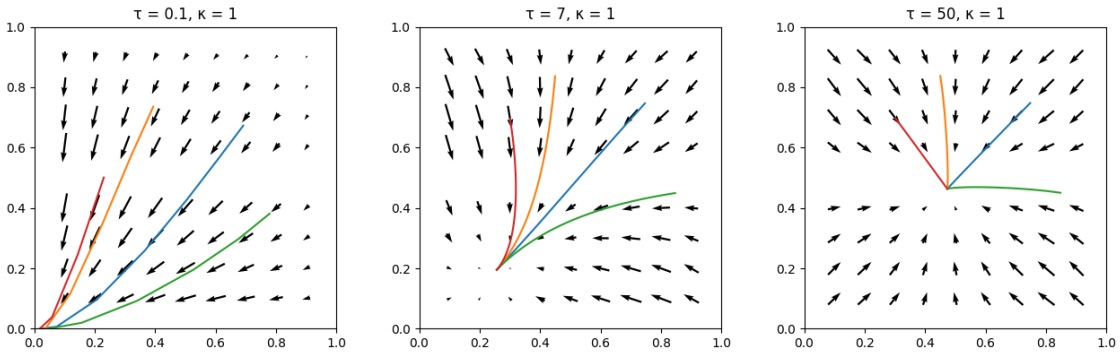

Figure 5: Influence of increasing τ on the Prisoner's Dilemma

4 Part 2 and 3

The task in parts 2 and 3 are similar: to train an agent to play both games. We will start of by applying some standard tabular learner to the problem, e.g. Cross Learning or Q-Learning. At some point (mostly for Leduc Poker), we will run into practical limits for tabular learners. This is where we explore approximation through neural networks. Deciding which features of the information state to use will be crucial. To improve performance, we will look into approximating hidden features (the opponent's cards). Teófilo et al. [9] can be a starting point for this part. When we run into the practical limits of these approximation methods, improving the efficiency of our learning architecture can be examined. Dulac-Arnold et al. [4] propose a tractable learning strategy that scales linearly in the number of actions. Their 'related work' section outlines some other strategies to improve the scalability of training reinforcement learning agents in the context of games.

References

[1] Daan Bloembergen. Analyzing reinforcement learning algorithms using evolutionary game theory. *Journal of theoretical biology*, 2010.

[2] Daan Bloembergen. *Multi-agent learning dynamics*. PhD thesis, 05 2015.

[3] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. Evolutionary dynamics of multi-agent learning:a survey. *Journal of Artificial Intelligence Research*, 2015.

[4] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces, 2015.

[5] Marc Harper, Vincent Knight, Martin Jones, Georgios Koutsovoulos, Nikoleta E. Glynatsi, and Owen Campbell. Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma. *PLOS ONE*, 12(12):1–33, 12 2017. doi: 10.1371/journal.pone.0188046. URL `https://doi.org/10.1371/journal.pone.0188046`.

[6] Marc Lanctot, Edward Lockhart, Jean-Baptiste Lespiau, Vinicius Zambaldi, Satyaki Upadhyay, Julien Pérolat, Sriram Srinivasan, Finbarr Timbers, Karl Tuyls, Shayegan Omidshafiei, Daniel Hennes, Dustin Morrill, Paul Muller, Timo Ewalds, Ryan Faulkner, János Kramár, Bart De Vylder, Brennan Saeta, James Bradbury, David Ding, Sebastian Borgeaud, Matthew Lai, Julian Schrittwieser, Thomas Anthony, Edward Hughes, Ivo Danihelka, and Jonah Ryan-Davis. Openspiel: A framework for reinforcement learning in games, 2019.

[7] Hisashi Ohtsuki and Martin A. Nowak. The replicator equation on graphs. *Journal of theoretical biology*, 243 1:86–97, 2006.

[8] Yoav Shoham and Kevin Leyton-Brown. *Multi-agent systems: Algorithmic, Game-Theoretic,and Logical Foundations*. 2009.

[9] Luís Filipe Teófilo, Nuno Passos, Luís Paulo Reis, and Henrique Lopes Cardoso. Adapting strategies to opponent models in incomplete information games: A reinforcement learning approach for poker. In Mohamed Kamel, Fakhri Karray, and Hani Hagras, editors, *Autonomous and Intelligent Systems*, pages 220–227, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31368-4.

[10] Karl Tuyls, Dries Heytens, Ann Nowe, and Bernard Manderick. Extended replicator dynamics as a key to reinforcement learning in multi-agent systems. In Nada Lavrač, Dragan Gamberger, Hendrik Blockeel, and Ljupčo Todorovski, editors, *Machine Learning: ECML 2003*, pages 421–431, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.