Derivation Write Up

Aamal Hussain

May 19, 2020

1 Introduction

- For use in Multi-Agent Reinforcement Learning
- Requirement to get some desired behaviours out of the system (e.g. conforming to predefined values or constraints)
- For this, it is required that we are able to predict the outcome of the learning behaviour
- Work done on determining the dynamics of MARL gives an indication of what solutions an algorithm will produce
- We can use this, alongside stability theory, to understand the nature of the games which result in stable behaviours and which result in chaotic behaviours.
- Largely follows the work presented by Sanders et al. However, whilst they consider EWA, we consider Q-Learning, which is more popular in RL communities.

1.1 Problem Statement

Given a particular learning algorithm and game, is it possible to determine whether the game is likely to reach a unique fixed point or exhibit more complex behaviour? What are the factors which affect this resulting behaviour?

1.2 Objectives and Scope

Analyse the stability of Multiagent Q-Learning on iterated games. Brings the work presented by Sanders et al to the reinforcement learning community.

Assumptions:

- Focuses on stateless games
- Homogeneous agents
- Discrete action spaces
- Small, finite number of agents.

2 Derivation

We start with the two-agent Q-Learning dynamics as presented by Tuyls et al.

$$\frac{\dot{x}(t)}{x(t)} = \alpha\tau(\sum_j a_{ij} y_j - \sum_{ij} x_i a_{ij} y_j) + \alpha \sum_j x_j ln(\frac{x_j}{x_i}) \tag{1a}$$

$$\frac{\dot{y}(t)}{y(t)} = \alpha\tau(\sum_j b_{ij} x_j - \sum_{ij} y_i b_{ij} x_j) + \alpha \sum_j y_j ln(\frac{y_j}{y_i}) \tag{1b}$$

In order to follow the conventions of spin glass theory for the analysis of disordered systems, we rescale the system so that the payoff matrix elements are of order $N^{-1/2}$.