Memory size in the Prisoner's Dilemma

Nikoleta E. Glynatsi   Vincent Knight

Abstract

The two player Iterated Prisoner's Dilemma is a fundamental iterated game used for studying the emergence of cooperation. The two players interact repeatedly and they have the ability to adopt strategies. A strategy allows a player to map the outcomes of the previous interactions to an action. A set of strategies that consider only the outcome of the previous round are called memory one. These players gain attention after a publication in 2012 that showed that a memory one strategy can manipulate its opponent.

In this manuscript we build upon a framework provided in 1989 for the study of these strategies and identify the best responses of memory one players. The aim of this work is to show the limitations of memory one strategies in multi-opponent interactions. A number of theoretic results are presented.

1   Introduction

The Prisoner's Dilemma (PD) is a two player person game used in understanding the evolution of co-operative behaviour. Each player can choose between cooperation (C) or defection (D). The decisions are made simultaneously and independently. The normal form representation of the game is given by:

$$S_p = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \quad S_q = \begin{pmatrix} R & T \\ S & P \end{pmatrix} \tag{1}$$

where S_p represents the utilities of the first player and S_q the utilities of the second player. The payoffs, (R, P, S, T), are constrained by equations (2) and (3). Constrain (2) ensures that defection dominates cooperation and constrain (3) ensures that there is a dilemma. Because the sum of the utilities for both players is better when both choose cooperation. The most common values used in the literature are $(3, 1, 0, 5)$ [2].

$$T > R > P > S \tag{2}$$

$$2R > T + S \tag{3}$$

The PD is a one shot game, however it is commonly studied in a manner where the history of the interactions matters. The repeated form of the game is called the Iterated Prisoner's Dilemma (IPD) and in the 1980s following the work of [3, 4] it attracted the attention of the scientific community.

In [3] a computer tournament of the IPD was performed. A tournament is a series of rounds of the IPD between pairs of strategies. The topology commonly used, [3, 4], is that of a round robin where all contestants
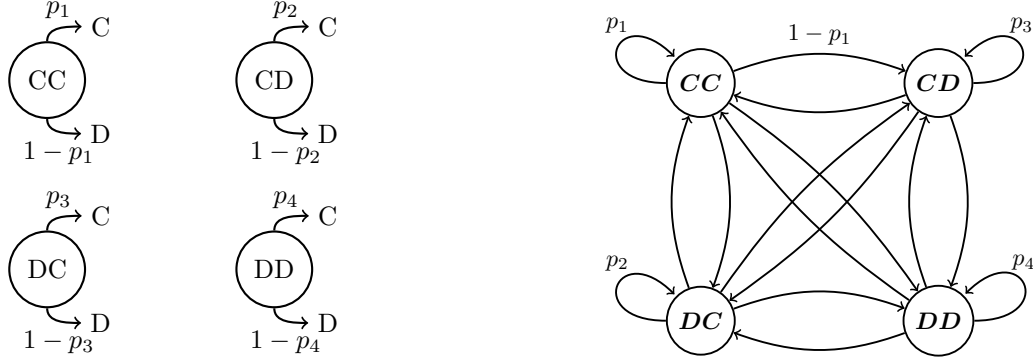
compete against each other. The winner of these tournament was decided on the average score and not in the number of wins.

These tournaments were the milestones of an era which to today is using computer tournaments to explore the robustness of strategies of IPD. Though the robustness can also be checked through evolutionary process [10]. However, this aspect will not be considered here, instead the focus is on performance in tournaments.

In Axelrod's original tournaments [3, 4], strategies were allowed access to the history and in the first tournament they also knew the number of total turns in each interaction. The history included the previous moves of both the player and the opponent. How many turns of history that a strategy would ue, the memory size, was left to the creator of the strategy to decide. For example the winning strategy of the first tournaments, Tit for Tat was a strategy that made use of the previous move of the opponent only. Tit for Tat is a strategy that starts by cooperating and then mimics the previous action of it's opponent. Strategies like Tit for Tat are called memory one strategies. A framework for studying memory one strategies was introduced in [8] and further used in [7, 9].

In [11] Press and Dyson, introduced a new set of memory one strategies called zero determinant (ZD) strategies. The ZD strategies, manage to force a linear relationship between the score of the strategy and the opponent. Press and Dyson, prove their concept of the ZD strategies and claim that a ZD strategy can outperform any given opponent.

The ZD strategies have tracked a lot of attention. It was stated that "Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma" [12]. In [12], the Axelrod's tournament have been re-run including ZD strategies and a new set of ZD strategies the Generous ZD. Even so, ZD and memory one strategies have also received criticism. In [6], the 'memory of a strategy does not matter' statement was questioned. A set of more complex strategies, strategies that take in account the entire history set of the game, were trained and proven to be more robust than ZD strategies.

1.1 Background

A memory one strategy p in a match against a strategy q would decide it's action in turn m based on what occurred in turn $m - 1$. If a strategy is concerned with only the outcome of a single turn then there are four possible 'states' the strategy could be in. These are CC, CD, DC, CC. A memory one strategy is denoted by the probabilities of cooperating after each of these states, $p = p_1, p_2, p_3, p_4 \in \mathbb{R}^4_{[0,1]}$

In 1989 a framework was introduced by to study the interactions of memory strategies.

- the mathematics - the state states - examples

As described in [9], a match between players p and q can be modelled as a stochastic process, where the players move from state to state. More specifically, it can be modelled by the use of a Markov process of four states. The transition probability matrix is defined as M and is given by,

$$M = \begin{bmatrix} p_1 q_1 & p_1(-q_1+1) & q_1(-p_1+1) & (-p_1+1)(-q_1+1) \\ p_2 q_3 & p_2(-q_3+1) & q_3(-p_2+1) & (-p_2+1)(-q_3+1) \\ p_3 q_2 & p_3(-q_2+1) & q_2(-p_3+1) & (-p_3+1)(-q_2+1) \\ p_4 q_4 & p_4(-q_4+1) & q_4(-p_4+1) & (-p_4+1)(-q_4+1) \end{bmatrix}. \tag{4}$$

Let v be the vector of the stationary probabilities of M and S_p payoff vector of player p. The states of vector v are given in the Appendix. The scores of each player can be retrieved by multiplying the probabilities of each state, at the stationary state, with the equivalent payoff. Thus, the utility for player p against q, denoted as $u_q(p)$, is defined by,

$$u_q(p) = v \times S_p. \tag{5}$$

1.2 Problem Description

The purpose of this work is to consider a given memory one strategy $p = (p_1, p_2, p_3, p_4)$, (in a similar fashion to [11]). However whilst [11] found a way for the opponent of p to manipulate q, this work will consider an optimisation approach to identify the best response $p^* = (p_1, p_2, p_3, p_4)$ to a strategy q. In essence answering the question: what is the best memory one strategy against a given memory one strategy.

This approach is then be expanded on to consider multiple players $q^{(1)}, q^{(2)}, \ldots, q^{(N)}$ which in turn allows for the optimisation of a given memory one strategy in a group of memory one strategies. Along the way a number of theoretic results will be proven:

- The utility of a given player p against a given opponent q is a ratio of quadratic forms.

- For given structural constraints on p ($p_1 = p_2 = p_3 = p_4$) a number of theoretic conditions are found to allow for further categorization of the utility of p.

- Optimization procedures, often reducing the complex optimisation problem to a search over a small finite set are found.

- Cases where defection is stationary.

2 Utility

Theorem 1 *For a given memory one strategy $p \in \mathbb{R}^4_{[0,1]}$ playing another memory one strategy $q \in \mathbb{R}^4_{[0,1]}$, the utility of the player $u_q(p)$ can be re written as a ratio of two quadratic forms:*

$$u_q(p) = \frac{\frac{1}{2}p^T Q p + c^T p + a}{\frac{1}{2}p^T \bar{Q} p + \bar{c}^T p + \bar{a}}, \tag{6}$$

where $Q, \bar{Q}$ are matrices of 4×4 defined with the transition probabilities of the opponent's transition probabilities q_1, q_2, q_3, q_4.

3

$$Q = \begin{bmatrix} 0 & -(q_1 - q_3)(q_2 - 5q_4 - 1) & q_3(q_1 - q_2) & -5q_3(q_1 - q_4) \\ -(q_1 - q_3)(q_2 - 5q_4 - 1) & 0 & (q_2 - q_3)(q_1 - 3q_4 - 1) & (q_3 - q_4)(5q_1 - 3q_2 - 2) \\ q_3(q_1 - q_2) & (q_2 - q_3)(q_1 - 3q_4 - 1) & 0 & 3q_3(q_2 - q_4) \\ -5q_3(q_1 - q_4) & (q_3 - q_4)(5q_1 - 3q_2 - 2) & 3q_3(q_2 - q_4) & 0 \end{bmatrix}, \quad (7)$$

$$\bar{Q} = \begin{bmatrix} 0 & -(q_1 - q_3)(q_2 - q_4 - 1) & (q_1 - q_2)(q_3 - q_4) & (q_1 - q_4)(q_2 - q_3 - 1) \\ -(q_1 - q_3)(q_2 - q_4 - 1) & 0 & (q_2 - q_3)(q_1 - q_4 - 1) & (q_1 - q_2)(q_3 - q_4) \\ (q_1 - q_2)(q_3 - q_4) & (q_2 - q_3)(q_1 - q_4 - 1) & 0 & -(q_2 - q_4)(q_1 - q_3 - 1) \\ (q_1 - q_4)(q_2 - q_3 - 1) & (q_1 - q_2)(q_3 - q_4) & -(q_2 - q_4)(q_1 - q_3 - 1) & 0 \end{bmatrix}. \quad (8)$$

c and $\bar{c}$, are 4×1 vectors defined by the transition rates q_1, q_2, q_3, q_4.

$$c = \begin{bmatrix} q_1(q_2 - 5q_4 - 1) \\ -(q_3 - 1)(q_2 - 5q_4 - 1) \\ -q_1 q_2 + q_2 q_3 + 3q_2 q_4 + q_2 - q_3 \\ 5q_1 q_4 - 3q_2 q_4 - 5q_3 q_4 + 5q_3 - 2q_4 \end{bmatrix}, \quad (9)$$

$$\bar{c} = \begin{bmatrix} q_1(q_2 - q_4 - 1) \\ -(q_3 - 1)(q_2 - q_4 - 1) \\ -q_1 q_2 + q_2 q_3 + q_2 - q_3 + q_4 \\ q_1 q_4 - q_2 - q_3 q_4 + q_3 - q_4 + 1 \end{bmatrix}. \quad (10)$$

Lastly, $a = -q_2 + 5q_4 + 1$ and $\bar{a} = -q_2 + q_4 + 1$.

2.1 Validation

All results are validated using [1] which is an open research framework for the study of the Iterated Prisoner's Dilemma. This package is described in [5].

To validate the formulation of $u_q(p)$ several memory one players were matched against 20 opponents. Note that the payoff values used hereafter in this work are $(3, 1, 0, 5)$. The simulated value of $u_q(p)$ has been calculated using [1] and the theoretical by substituting in equation (6). In Figure 2, both the simulated and the theoretical value of $u_q(p)$, against each opponent, are plotted for three different memory one strategies. Figure 2 indicates that the formulation of $u_q(p)$ as a quadratic ratio successfully captures the simulated behaviour.

3 Proximate of best responses

The analytical formulation gives the advantage of time. That is because the payoffs of a match between two opponents are now retrievable without simulating the actual match itself. Utility $u_q(p)$ can now be calculated numerically instead of using simulation. In the introduction a question was raised: which memory one strategy is the **best response** against another memory one? This will be considered as an optimisation problem, where a memory one strategy p wants to optimise it's utility $u_q(p)$ against an opponent q. The decision variable is the vector p and the solitary constrains $p \in \mathbb{R}^4_{[0,1]}$. The optimisation problem is given by (11).

$$\max_p : \quad \frac{\frac{1}{2}pQp^T + c^T p + a}{\frac{1}{2}p\bar{Q}p^T + \bar{c}^T p + \bar{a}}$$
$$\text{such that :} \quad p \in \mathbb{R}^4_{[0,1]}. \quad (11)$$
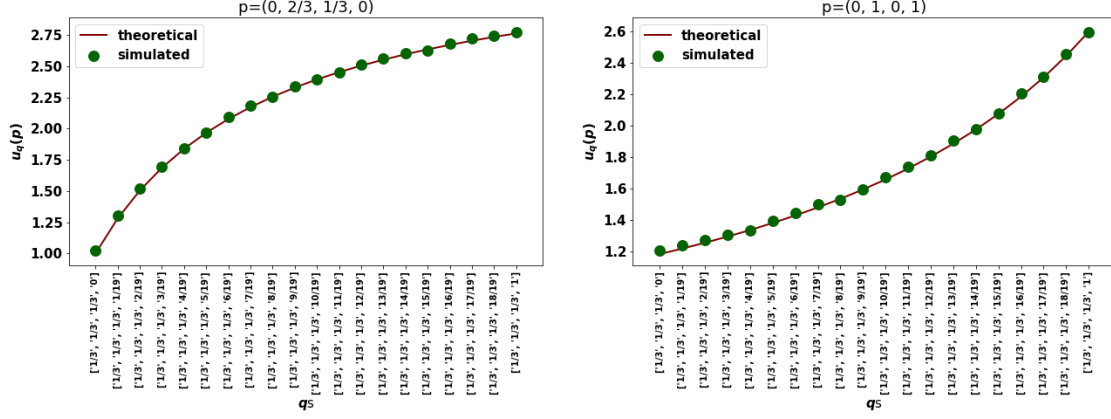
Figure 2: Differences between simulated and analytical results.

3.1 In purely random strategies

In this section we explore a simplified version of the main problem. A set of memory one strategies where the transition probabilities of each state are the same, are called **purely random strategies**. as

The optimisation problem of (11) now has an extra constraint and is re written as,

$$
\begin{aligned}
\max_p &: u_q(p) \\
\text{such that} &: p_1 = p_2 = p_3 = p_4 = p \\
& 0 \leq p \leq 1.
\end{aligned}
\tag{12}
$$

3.2 In memory one strategies

4 Stability of defection

5 Numerical experiments

References

[1] The Axelrod project developers . Axelrod: ¡release title¿, April 2016.

[2] R Axelrod and WD Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[3] Robert Axelrod. Effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(1):3–25, 1980.

[4] Robert Axelrod. More effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(3):379–403, 1980.

[5] Vincent Knight, Owen Campbell, Marc Harper, Karol Langner, James Campbell, Thomas Campbell, Alex Carney, Martin Chorley, Cameron Davidson-Pilon, Kristian Glass, Tomáš Ehrlich, Martin Jones, Georgios Koutsovoulos, Holly Tibble, Müller Jochen, Geraint Palmer, Paul Slavin, Timothy Standen, Luis Visintini, and Karl Molden. An open reproducible framework for the study of the iterated prisoner's dilemma. 1(1), 2016.

[6] Christopher Lee, Marc Harper, and Dashiell Fryer. The art of war: Beyond memory-one strategies in population games. *PLOS ONE*, 10(3):1–16, 03 2015.

[7] Frederick A Matsen and Martin A Nowak. Win–stay, lose–shift in language learning from peers. *Proceedings of the National Academy of Sciences*, 101(52):18053–18057, 2004.

[8] Martin Nowak and Karl Sigmund. Game-dynamical aspects of the prisoner's dilemma. *Applied Mathematics and Computation*, 30(3):191–213, 1989.

[9] Martin Nowak and Karl Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20(3):247–265, 1990.

[10] Martin A Nowak. *Evolutionary Dynamics: Exploring the Equations of Life.* Cambridge: Harvard University Press.

[11] William H. Press and Freeman J. Dyson. Iterated prisoners dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.

[12] Alexander J. Stewart and Joshua B. Plotkin. Extortion and cooperation in the prisoners dilemma. *Proceedings of the National Academy of Sciences*, 109(26):10134–10135, 2012.