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.

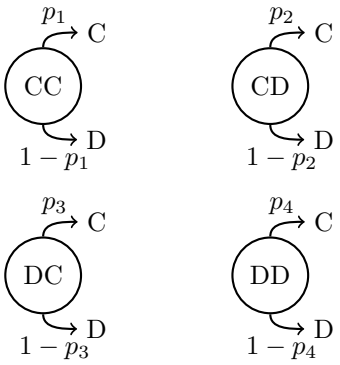
2 Problem Description

The work of [11] is considered a milestone in the study of the IPD. The authors managed to discover a new family of memory on strategies that are able to manipulate another player. Their work however suffered from limitations. Only pair wise interactions are taken into account and not multi agent interactions; interactions in a tournament setting.

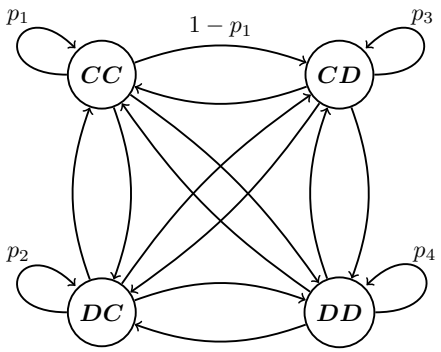
This manuscript will consider an optimisation approach to identify the best response memory one strategy. That will be done for a given opponent as well as in tournaments. The aim of this work in to:

- construct compact methods of identifying best responses.
- numerically tests the analytical methods.
- prove that memory one strategies suffer for their memory size constrain.

The search of best responses will be considered as an optimisation problem. In the following section the formulation is introduced as well as the objective function.



(a) Diagrammatic representation of a memory one strategy.



(b) Markov chain.

2.1 Utility

The objective function in a match of the IPD can be no other than the utility the players receive at the end of the match. More specifically we will consider the utility of a given memory one strategy against another such strategy.

Initially we properly define memory one strategies. A memory one strategy p in a match against a strategy q would decide it's next action in turn based on only what occurred in the previous turn. 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, $p = p_1, p_2, p_3, p_4 \in \mathbb{R}^4_{[0,1]}$ as shown by Figure 1a.

In [9] a framework was introduced by to study the interactions of memory strategies as a stochastic process. It was described how 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, shown in Figure 1b.

The transition probability matrix is defined as M and is given by,

$$M = \begin{bmatrix} p_1q_1 & p_1\left(-q_1+1\right) & q_1\left(-p_1+1\right) & \left(-p_1+1\right)\left(-q_1+1\right) \\ p_2q_3 & p_2\left(-q_3+1\right) & q_3\left(-p_2+1\right) & \left(-p_2+1\right)\left(-q_3+1\right) \\ p_3q_2 & p_3\left(-q_2+1\right) & q_2\left(-p_3+1\right) & \left(-p_3+1\right)\left(-q_2+1\right) \\ p_4q_4 & p_4\left(-q_4+1\right) & q_4\left(-p_4+1\right) & \left(-p_4+1\right)\left(-q_4+1\right) \end{bmatrix}. \tag{4}$$

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,

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

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.

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

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

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

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

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

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

2.1.1 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 Best responses Analytically

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?

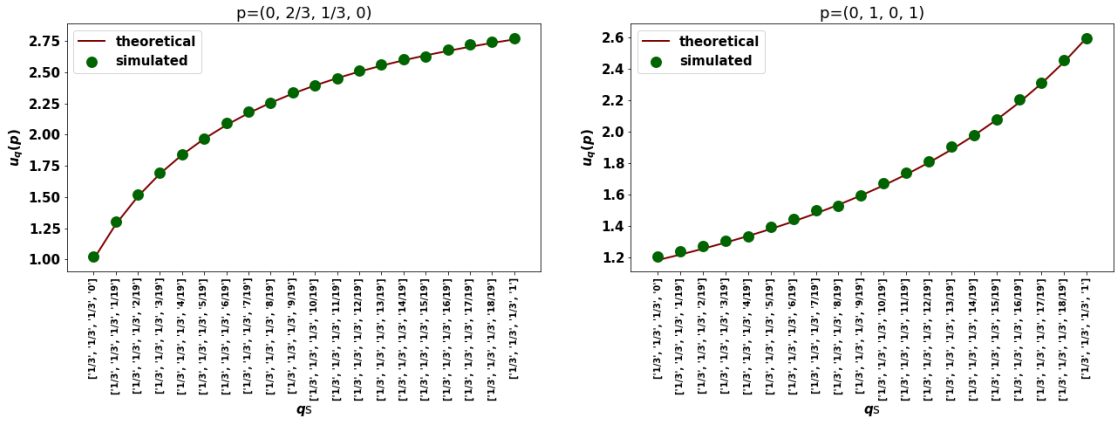


Figure 2: Differences between simulated and analytical results.

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).

$$\begin{aligned} \max_p : \quad & \frac{\frac{1}{2}pQp^T + c^T p + a}{\frac{1}{2}p\bar{Q}p^T + \bar{c}^T p + \bar{a}} \\ \text{such that} : \quad & p \in \mathbb{R}^4_{[0,1]}. \end{aligned} \tag{11}$$

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,

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

Note that now $u_q(p)$ is given by,

$$\begin{aligned} u_q(p) = \frac{n_2 p^2 + n_1 p + n_0}{d_1 p + d_0}, \text{ where} \\ n_2 = -q_1 + q_2 + 2q_3 - 2q_4, \ n_1 = q_1 - 2q_2 - 5q_3 + 7q_4 + 1, \ n_0 = q_2 - 5q_4 - 1 \text{ and} \\ d_1 = q_1 - q_2 - q_3 + q_4, \ d_0 = q_2 - q_4 - 1. \end{aligned} \tag{13}$$

The utility now is a function of a single variable. Moreover due the format of the utility we know that the maximum degree is that of a degree equal to 2. Thus for identifying the best response the following theorem

is stated:

Theorem 2 (Optimisation of purely random player in a match) *The optimal behaviour of a* ***purely random*** *player* (p,p,p,p) *against a memory one opponent* q *is given by:*

$$p^* = \text{argmax}(u_q(p)),\ p \in S_q,$$

where the set S_q *is defined as*

$$S_q = \left\{ 0, p_\pm, 1 \;\middle|\; \begin{array}{l} 0 < p_\pm < 1, \\ p_\pm \neq \frac{-d_0}{d_1} \end{array} \right\}$$

Moreover, due the form of the utility for the purely random case, a few interesting cases have been identified. These are:

- Constant case. The plater p becomes indifferent
- Linear case. The utility against q is linear. Thus the best response is a pure strategy.

For each case the following Lemmas are stated equivalently. Proofs can be found in the appendix.

Lemma 3 *A given memory one player,* (q_1,q_2,q_3,q_4)*, makes a* ***purely random*** *player,* (p,p,p,p)*, indifferent if and only if,* $-q_1+q_2+2q_3-2q_4 = 0$ *and* $(q_2-q_4-1)(q_1-2q_2-5q_3+7q_4+1)-(q_2-5q_4-1)(q_1-q_2-q_3+q_4) = 0$.

Lemma 4 *Against a memory one player,* (q_1,q_2,q_3,q_4)*, a* ***purely random*** *player would always play a pure strategy if and only if* $(q_1q_4 - q_2q_3 + q_3 - q_4)(4q_1 - 3q_2 - 4q_3 + 3q_4 - 1) = 0$.

Now in a tournament the utility is changed and can be written as:

$$\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) = \frac{1}{N} \frac{\sum_{i=1}^{N}(n_2^{(i)}p^2 + n_1^{(i)}p + n_0^{(i)}) \prod_{\substack{j=1 \\ j\neq i}}^{N} (d_1^{(j)}p + d_0^{(j)})}{\prod_{i=1}^{N}(d_1^{(i)}p + d_0^{(i)})}. \tag{14}$$

Let us ignore the coefficients of the utility for now and only consider the degree of $u_{q(i)}(p)$. Note that the degree of $\sum_{i=1}^{N}(n_2^{(i)}p^2 + n_1^{(i)}p + n_0^{(i)})$ is N .The sum being multiplied by the product of $\prod_{\substack{j=1 \\ j\neq i}}^{N} (d_1^{(j)}p + d_0^{(j)})$ with a degree N. Thus the numerator is determined to be a polynomial of degree $N+1$. The denominator is also a polynomial of degree N. Now that the degree have been established, equation (14) can be simplified to the following ratio of polynomials:

$$\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) = \frac{1}{N}\left(\frac{\sum_{i=0}^{N+1} h_i p^i}{\sum_{i=0}^{N} k_i p^i}\right) \tag{15}$$

Following a similar approach to the study of matches, see equation (**??**), now $\frac{d{u_q}_{(i)}(p)}{dp}$ can be calculated. Differentiating equation (14) yields to the following:

Similar to identifying p^* in match, the roots of $\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p)$ alongside the bounds compose the candidate set of solutions. The roots of $\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p)$ are the roots only of the numerator, as the denominator can not be nullified. Studying equation (**??**), the degree of the numerator can be verified to be equal to $2N$. Thus, the size of roots of the numerator is equal to $2N$.

The roots on the polynomial in this work will be calculated using a companion matrix method [**?**]. This method allows the roots of the polynomial to be computed by calculating the eigenvalues of the corresponding companion matrix.

Corresponding to Theorem 2, a theorem for p^* in a tournament is given by Theorem 5.

Theorem 5 (Optimisation of purely random player in a tournament) *The optimal behaviour of a* ***purely random*** *player (p, p, p, p) in an $N-$memory one player tournament, $\{q_{(1)}, q_{(2)} \ldots, q_{(N)}\}$ is given by:*

$$p^* = \text{argmax}(\sum_{i=1}^{N} {u_q}^{(i)}(p)),\ p \in S_{q(i)},$$

where the set S_q is defined as:

$$S_q = \underset{\lambda_i \neq \frac{do_i}{d1_i}}{U_{i=1}^{2N}} \lambda_i \cup \{0, 1\},$$

and λ_i are the eigenvalues of the companion matrix corresponding to the numerator of

$$\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p).$$

Note that the size of candidate solutions is $1 \leq |S_{q(i)}| \leq 2N + 2$.

3.2 In memory one strategies

- Optimisation in matches.
- + Lagrange
- Optimisation in tournaments.

4 Best responses Numerically

- Purely random in matches. Yes our algorithm works
- Purely random in tournaments. Yes our algorithm works
- Reactive strategies and resultant theory
- Differential evolution

5 Limitation of memory

- Purely random in tournaments, compare with complex
- Parameter sweep and how gambler do better?

6 Extra: Stability of defection

- Stability of defection against mem one
- Stability of defection against reactive

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.