Memory size in the Prisoner's Dilemma

Nikoleta E. Glynatsi Vincent Knight

Abstract

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) and 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 $(3, 1, 0, 5)$ [4].

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

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

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 [5, 6] it attracted the attention of the scientific community.

In [5] and [6], the first well known computer tournaments of the IPD were performed. A total of 13 and 63 strategies were submitted in computer code and competed against each other in a round robin tournament. All contestants competed against each other, themselves and random strategy and the winner was decided on the average score a strategy achieved and not in the number of wins. The strategies were allowed access to the full history of each match. The history included the previous moves of both the player and the opponent. How many turns of history that a strategy would use, the memory size, was left to the creator of the strategy to decide.

The winning strategy of both tournaments was a strategy called Tit for Tat. Tit for Tat is a strategy which starts by cooperating and then mimics the last move of it's opponent. This is a strategy which makes use of the previous move of the opponent only and it's called a reactive strategy. Reactive strategies have been used to explore best behaviour in the IPD and a framework for studying such strategies was introduced in [15]. This was later used to introduce other reactive strategies such as Generous Tit For Tat [16].

Reactive strategies are a subset of memory one strategies. Memory one strategies similarly are only concern with the previous turn. However, they take into consideration both players' recent moves to decide on an action. Several successful memory one strategies are found in literature, for example Pavlov [14].

A well known set of memory on strategies were introduced in [17] called zero determinant (ZD) strategies. The ZD strategies, manage to force a linear relationship between the score of the strategy and the opponent. The authors showed that ZD strategies were a set of strategies that could dominate any evolutionary opponent, in pairwise interactions. ZD strategies were dominating using a single size memory and authors questioned the usefulness of memory in the IPD.

The ZD strategies tracked a lot of attention. It was stated that "Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma" [18]. In [18], 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 [13], 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 The Problem

In this manuscript we explore the size of memory for strategies of IPD.

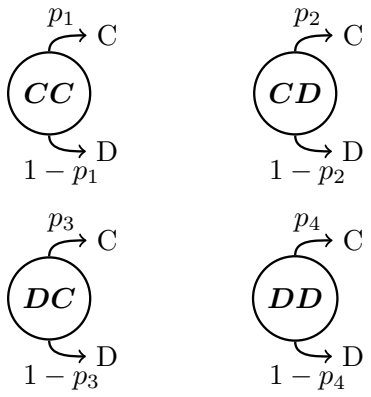
The purpose of this work is to consider a given memory one strategy in a similar fashion to [17]. However whilst [17] found a way for a player to manipulate an opponent, this work will consider an optimisation approach to identify the best response to that opponent. In essence the aim is to produce a compact method of identifying the best memory one strategy against a given opponent.

In the second part of this manuscript we explore the limitation of these best response strategies. This is achieved by comparing the performance of an optimal memory one strategy, for a given environment, with the performance of a more complex strategy. The type of complex strategy used is described in depth in the following Sections. Keep in mind, that the complex strategies used here has a memory greater than one and it keeps track of the initial move.

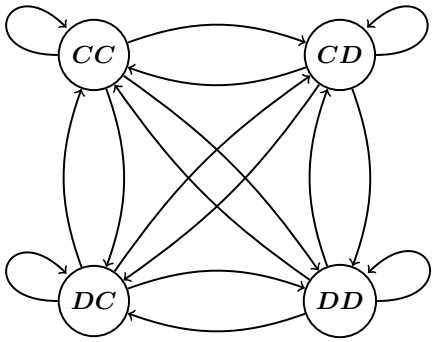
2 Methodology

In [15] a framework was introduced to study the interactions of memory one strategies modelled as a stochastic process. The work [17] of is also build upon at manuscript. There it is stated that 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, $p = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4_{[0,1]}$. A diagrammatic representation of such strategy is given in Figure 1a.

Moreover, if two players are moving from state to state using the transition probabilities, this can be modelled as a Markov process of four states, shown by Figure 1b. The transition matrix M of Figure 1b is defined as,



(a) Diagrammatic representation of a memory one strategy.



(b) Markov chain on a PD game.

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

where the vector of the stationary probabilities of is v(given in the Appendix). Combining the stationary vector v with the payoff matrices1 allow us to retrieve the expected outcome for each player. Thus, the utility for player p against q, denoted as $u_q(p)$, can defined by,

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

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. This analytical formulation will be used hereupon and it will will allow us to study best responses in an analytical way.

Though this formulation, which was described in 1989 [15], have been used in several different works not many insights have been given for form of $u_q(p)$. Here we present one of our first results which concerns the form $u_q(p)$. That is that $u_q(p)$ is given by a ratio of two quadratic forms, as presented by Theorem 1.

Theorem 1 *The expected utility of a memory one strategy $p \in \mathbb{R}^4_{[0,1]}$ against a memory one opponent strategy $q \in \mathbb{R}^4_{[0,1]}$, denoted as $u_q(p)$, can be written as a ratio of two quadratic forms:*

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

where $Q, \bar{Q}$ 4×4 matrices defined by the transition probabilities of the opponent q_1, q_2, q_3, q_4 as follows:

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

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

c and $\bar{c}$, are 4×1 vectors defined by:

$$c = \begin{bmatrix} q_1(q_2 - 5q_4 - 1) \\ -(q_3 - 1)(q_2 - 5q_4 - 1) \\ -q_1q_2 + q_2q_3 + 3q_2q_4 + q_2 - q_3 \\ 5q_1q_4 - 3q_2q_4 - 5q_3q_4 + 5q_3 - 2q_4 \end{bmatrix}, \tag{9}$$

$$\bar{c} = \begin{bmatrix} q_1(q_2 - q_4 - 1) \\ -(q_3 - 1)(q_2 - q_4 - 1) \\ -q_1q_2 + q_2q_3 + q_2 - q_3 + q_4 \\ q_1q_4 - q_2 - q_3q_4 + q_3 - q_4 + 1 \end{bmatrix}. \tag{10}$$

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

The formulation of $u_q(p)$ has been validated using numerical experiments. Several memory one players were matched against 20 opponents and their theoretical utility $u_q(p)$ was compared to a simulated one. Figure 2 indicates that the formulation of $u_q(p)$ as a quadratic ratio successfully captures the simulated behaviour.

The simulated utility, denoted as $U_q(p)$ has been calculated using [1]. It is an open research framework for the study of the IPD and is described in [12]. Note that when referring to $U_q(p)$ here onwards we mean the simulated utility calculated with [1].

Now that the analytical formulation has been validated in the following Sections is used to explore bests responses in memory one strategies. Moreover, the new formulation of Theorem 1 allow us to retrieve a number of theoretical results. These are also discussed in the next sections.

3 Analytically Results

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).

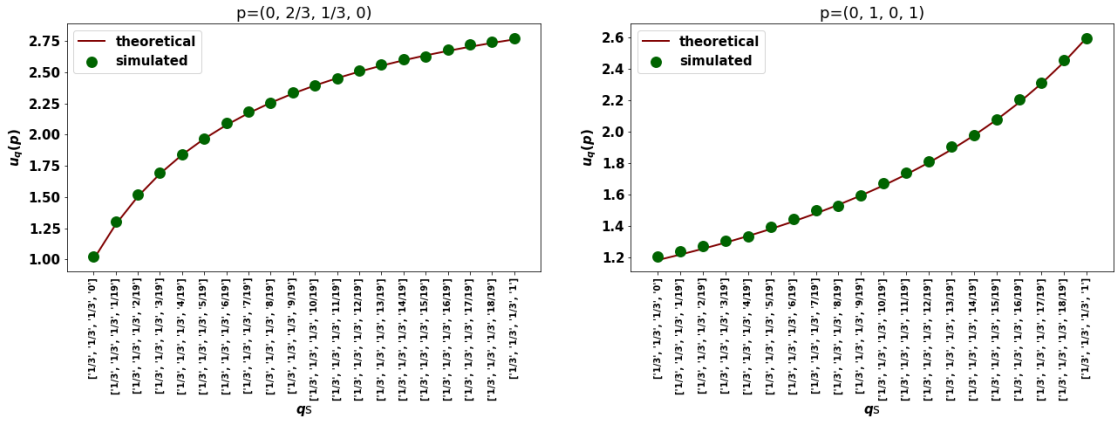


Figure 2: Differences between simulated and analytical results.

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

3.1 Convexity

This work is concerned with a fractional optimisation problem of quadratic forms. Initially, the convexity, whether or not $u_q(p)$ is concave [10], is checked (concave because is a maximisation problem).

To test the hypothesis that $u_q(p)$ is concave an empirical analysis was performed using computer code. It was shown that there exists at least one point for which the definition of concavity does not hold.

Several articles in fractional optimisation of quadratic forms that were non concave can be found [7, 8]. Though in these works both the numerator and denominator of the fractional problem were concave. In [3] it is stated that a quadratic form will be concave if and only if it's symmetric matrix is negative semi definite. In Appendix, it is proved that neither the numerator or the denominator of equation (6) are concave.

3.2 Best responses

The non concavity of $u(p)$ indicates multiple local optimal points. Thus we are not searching a single optimal point but a set of candidate optimal points. The aim is to introduce a compact way of constructing the candidate set. Once the set is defined the point that maximises $u(p)$ corresponds to the best response strategy.

The problem considered is a bounded because $p \in \mathbb{R}^4_{[0,1]}$. Because of this it is known that the candidate solutions will exist either at the boundaries of the feasible solution space, either within that space. The method of Lagrange Multipliers and Karush-Kuhn-Tucker conditions also agrees with this conclusion. The Karush-Kuhn-Tucker conditions are used because our constrains are inequalities.

Thus the candidate solution set is constructed as follows:

- any or all of p_1, p_2, p_3, p_4 are $\in 0, 1$
- the rest or all of p_1, p_2, p_3, p_4 are given by the roots of $\frac{du}{dp}$.

The derivative $\frac{du}{dp}$ is given by,

$$\begin{aligned}\frac{du}{dp} = & \frac{(\frac{1}{2}pQp^T + cp + a)'(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}) - (\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})'(\frac{1}{2}pQp^T + cp + a)}{(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})^2} \\ = & \frac{(pQ + c)(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}) - (p\bar{Q} + \bar{c})(\frac{1}{2}pQp^T + cp + a)}{(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a})^2}\end{aligned} \tag{12}$$

For equation 12 to be zero, the numerator must fall to zero and the denominator can not nullified. Thus we conclude that the best response of a memory one strategy in match is given by Lemma 2.

Lemma 2 *The optimal behaviour of a memory one strategy player (p^*) against a given opponent q is given by:*

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

where the set S_q is defined as

$$S_q = \{0, \bar{p}, 1\}^4$$

where the vector $\bar{p}$ is the vector for which the following conditions are true:

$$(pQ + c)(\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a}) - (p\bar{Q} + \bar{c})(\frac{1}{2}pQp^T + cp + a) = 0 \tag{13}$$

and

$$\frac{1}{2}p\bar{Q}p^T + \bar{c}p + \bar{a} \neq 0 \tag{14}$$

Note that equation 13 is a 4− polynomial system of 4 variables. Each polynomial corresponds to a partial derivative of $u_q(p)$.

A question that arises immediately after capturing best responses of memory one strategies in pairwise interactions is: What is the optimal memory player against multiple opponents, in a tournament environment. Let us consider a collection of opponents: $\{q^{(1)}, q^{(2)}, \ldots, q^{(N)}\}$, finding the optimal behaviour is captured as:

$$\max_p : \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p)$$
$$st : p \in \mathbb{R}_{[0,1]} \tag{15}$$

where,

$$\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) = \frac{1}{N}\frac{\sum_{i=1}^{N}(\frac{1}{2}pQ^{(i)}p^T + c^{(i)}p + a^{(i)}) \prod_{\substack{j=1 \\ j\neq i}}^{N}(\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}{\prod_{i=1}^{N}(\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)}p + \bar{a}^{(i)})}. \tag{16}$$

Thus, we are optimising against the average utility over the set of opponents. Note that the best response can not be captured by optimising against the mean opponent. Thus,

$$\max_p \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) \neq \max_p u_{\frac{1}{N}\sum_{i=1}^{N} q^{(i)}}(p). \tag{17}$$

A number of numerical experiments have been performed for cases where $p = (p, p, p, p$ and $p = (p_1, p_2, p_1, p_2)$. This was done in order to compare the right hand side of equation (17) to the left. The fact that equation (17) holds is evident by Figure 3.

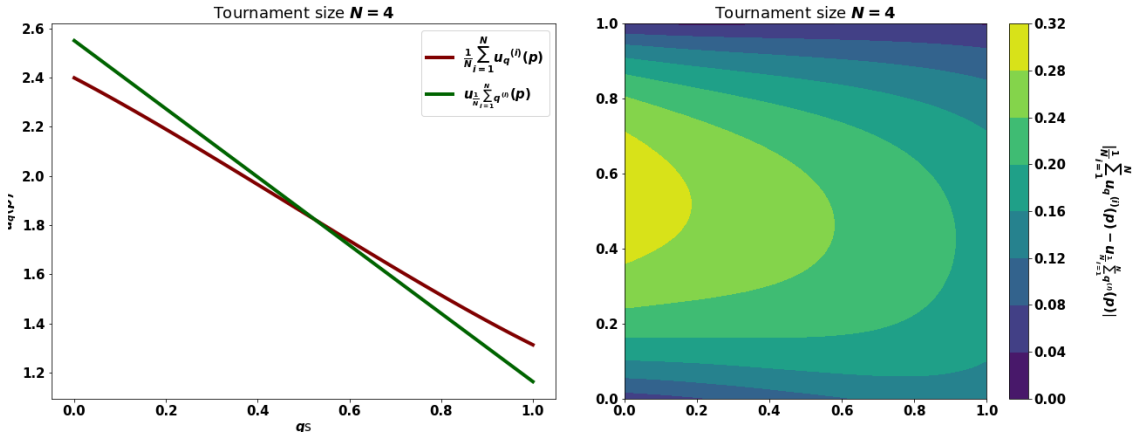


Figure 3: Hypothesis.

A similar approach to the one considered for problem 11 needs to be used for problem 15 as well. The candidate solutions set would be constructed by considering the bounds of the feasible space and the roots of the derivative $\frac{d}{dp}\frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p)$. The derivative is given by,

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

$$= \frac{(\sum_{i=1}^{N} {Q_N^{(i)}}' \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)} + \sum_{i=1}^{N} {Q_D^{(i)}}' \sum_{\substack{j=1 \\ j\neq i}}^{N} Q_N^{(i)} \prod_{\substack{l=1 \\ l\neq i \\ l\neq j}}^{N} Q_D^{(i)}) \times \prod_{i=1}^{N} Q_D^{(i)} - (\sum_{i=1}^{N} {Q_D^{(i)}}' \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)}) \times (\sum_{i=1}^{N} Q_N^{(i)} \prod_{\substack{j=1 \\ j\neq i}}^{N} Q_D^{(i)})}{(\prod_{i=1}^{N} Q_D^{(i)})^2} \tag{18}$$

where,

$$\begin{aligned} Q_N^{(i)} &= \frac{1}{2} p Q^{(i)} p^T + c^{(i)} p + a^{(i)}, \\ {Q_N^{(i)}}' &= p Q^{(i)} + c^{(i)}, \\ Q_D^{(i)} &= \frac{1}{2} p \bar{Q}^{(i)} p^T + \bar{c}^{(i)} p + \bar{a}^{(i)}, \\ {Q_D^{(i)}}' &= p \bar{Q}^{(i)} + \bar{c}^{(i)}. \end{aligned}$$

Extracting the roots of equation's (18) numerator is not an easy task. This also applied to equation (13). Both equations are a system of 4 polynomials and the degree of the polynomials is gradually increasing every time an extra opponent is taken into account.

Because of that no further analytical consideration is given to problems 11 and 15. Instead both best responses of memory one strategies, pairwise and in multi interactions, will be solved using numerical methods. The methods are their results are discussed in the following Section.

Though best responses can no longer be explored in an exact analytical way there are still many advantages by the formulation of Theorem 1. In the following subsections several theoretical results and exact ways of identifying best responses in constrained versions of problem 15 are presented. Moreover, the robustness of defection in specific interactions is investigated.

3.3 Purely random

The first constrained problem to be explored is that of the purely random strategies. Purely random strategies are a set of memory one strategies where the transition probabilities of each state are the same. The optimisation problem of (11) now has an extra constraint and is re written as,

$$\begin{aligned} \max_p : & \frac{1}{N}\sum_{i=1}^{N} {u_q}^{(i)}(p) \\ \text{such that} : & 0 \leq p \leq 1 \\ & p_1 = p_2 = p_3 = p_4 = p. \end{aligned} \tag{19}$$

Due to the additional constrain, $\sum_{i=1}^{N} {u_q}^{(i)}(p)$ is now a function of a single variable p and it can be handled analytically. To construct the set of candidate solutions a similar approach as the one described in previous sections is used. Thus:

- either p is $\in 0, 1$
- or p is given by the roots of $\frac{d\sum_{i=1}^{N} {u_q}^{(i)}(p)}{dp}$.

The roots of the derivative are given by nullifying the numerator of the derivative. It has been proved, Appendix, that the degree of the numerator does not exceed $2N$ where N is the number of opponents. Thus there are $2N$ possible roots in the feasible space. These results are summarized by Lemma 3.

Lemma 3 (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:

$$S_q = \{0, \lambda_i, 1\},\ \textit{for } i \in [1, 2N].$$

Note that λ_i are the eigenvalues of the companion matrix corresponding to the numerator of

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

and λ_i are such that the denominator is not nullified.

To compute the roots of the polynomial the algorithm used is that of computing the eigenvalues of the corresponding companion matrix. The algorithm is presented and discussed in [9].

Furthermore, for the case of the purely random players two more theoretical results are discussed. These are the cases where:

- the opponent has manage to make a random player indifferent
- the best behaviour is a pure strategy.

There is importance in both results. Initially, being indifferent refers to our actions no having any effects on the match. Thus our behaviour can not be optimised. Secondly, by a pure strategy we are referring to the edge cases, behaving as a defector or a cooperator. In in those case it is know that p^* is $\in 0, 1$. The results are given equivalently by Lemmas 4 and 5 and they are respective to the actions of the opponent.

Lemma 4 *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 5 *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* $(q_1q_4 - q_2q_3 + q_3 - q_4)(4q_1 - 3q_2 - 4q_3 + 3q_4 - 1) = 0$.

Figure 4 illustrates that conditions given by Lemma 4 and 5 capture the behaviour as described.

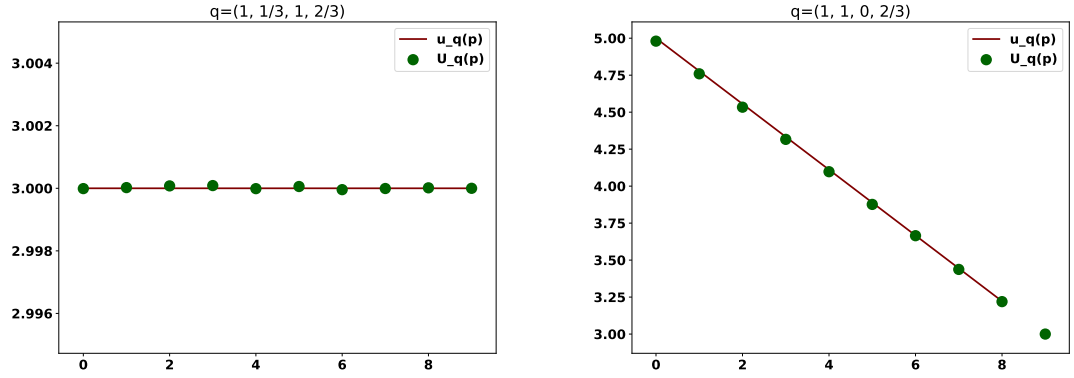


Figure 4: Proof of concept for Lemmas 4, 5.

3.4 Reactive Strategies

The next constrained case considered here is that of the reactive strategies. Reactive strategies are a set of memory one strategies where they only take into account the opponent's previous moves. As described in Section 1 Tit for Tat is a reactive strategy. The optimisation problem of (11) now has an extra constraint and is re written as,

$$\begin{aligned} \max_p &: u_q(p) \\ \text{such that} &: p_1 = p_3 \text{ and } p_2 = p_4 \\ &0 \leq p_1, p_2 \leq 1. \end{aligned} \tag{20}$$

Reactive strategies allow us to study u_p as a function of two variables p_1, p_2. The candidate solution set can be constructed as follows:

- p_1, p_2 are $\in \{0, 1\}^2$
- the rest or all of p_1, p_2 are given by the roots of $\frac{du}{dp}$.

Note that now,

$$\frac{du}{dp} = 0$$

is now a system of 2 polynomials over 2 variables. Each polynomial is equivalent to a partial derivative over p_1 and p_2. There are many methods that allow us to solve that analytical. In this work we will be using resultant theory to extract the roots.

The resultant is a symmetric function of the roots of the polynomials of a system and it can be expressed as a polynomial in the coefficients of the polynomials. The resultant will equal zero if and only if the system has at least one common root. Thus, the resultant becomes very useful in identifying whether common roots exist.

In this work the Sylvester's resultant [2] denoted as (R_S) is considered. The Sylvester's resultant is used to solve system of a single variable. However, for a system of two variables we solve over one variable and the second is kept as a coefficient. Thus we can find the roots of the equations and that is why the resultant is often refereed to as the eliminator.

Thus the best response of a reactive strategy is given in very similar approach as that described in Lemma 11. However now the partial derivatives can be solved using an exact method. Note that for pairwise interactions the maximum degree of the polynomials is equal to $2N$, however the degree increases as opponents are introduced.

3.5 Stability of defection

The final theoretical result explored is the stability of defection. Defection is the dominant strategy in the one shot game and it can be proven to be an optimal strategy in given environments.

In this manuscript we try to provide a condition for when defection is the best response. This will be done by considering equation (12). Let equation (12) for $p_0 = (0, 0, 0, 0)$ given by,

$$\frac{du}{dp_0} = \frac{c\bar{a} - \bar{c}a}{\bar{a}^2}. \tag{21}$$

The numerator $\bar{c}a - c\bar{a}$ is given by,

$$\begin{bmatrix} 0 \\ 0 \\ q_4\left(4q_1q_2 - 3q_2^2 - 4q_2q_3 + 3q_2q_4 + 4q_3 - 5q_4 - 1\right) \\ -\left(-q_2 + q_4 + 1\right)\left(-5q_1q_4 + 3q_2q_4 + 5q_3q_4 - 5q_3 + 2q_4\right) - \left(-q_2 + 5q_4 + 1\right)\left(q_1q_4 - q_2 - q_3q_4 + q_3 - q_4 + 1\right) \end{bmatrix}$$

and $\bar{a}^2 = (-q_2 + q_4 + 1)^2$ which is always positive. In order for defection to be the best response the derivative must have a negative sign at the point p_0. That means that the utility is only decreasing after p_0 and the points before are outside our feasible space.

The sign of the derivative is given by the numerator, thus $\bar{c}a - c\bar{a}$. More specifically from equations,

$$q_4\left(4q_1q_2 - 3q_2^2 - 4q_2q_3 + 3q_2q_4 + 4q_3 - 5q_4 - 1\right) \tag{22}$$

$$-\left(-q_2 + q_4 + 1\right)\left(-5q_1q_4 + 3q_2q_4 + 5q_3q_4 - 5q_3 + 2q_4\right) - \left(-q_2 + 5q_4 + 1\right)\left(q_1q_4 - q_2 - q_3q_4 + q_3 - q_4 + 1\right) \tag{23}$$

Both signs of the partial derivatives must be negative in order for the overall function to be decreasing, thus defection being the best response. The signs of equations (22) and (23) vary. There are cases that they have the same sign and cases that they do not, this is shown by numerical example summarized in Table 1.

					equation(22)	equation(23)
1	$q_1 = \frac{3}{10}$,	$q_2 = \frac{3}{20}$,	$q_3 = \frac{13}{20}$,	$q_4 = \frac{7}{100}$	+	+
2	$q_1 = \frac{11}{25}$,	$q_2 = \frac{3}{10}$,	$q_3 = \frac{9}{10}$,	$q_4 = \frac{1}{2}$	-	-
3	$q_1 = \frac{17}{20}$,	$q_2 = \frac{3}{4}$,	$q_3 = \frac{2}{5}$,	$q_4 = \frac{1}{4}$	-	+
4	$q_1 = \frac{13}{88}$,	$q_2 = \frac{21}{92}$,	$q_3 = \frac{21}{26}$,	$q_4 = \frac{20}{67}$	+	-

Table 1: Numerical examples of the derivative's sign.

Lets us consider a constrained version of the problem once again. Lets us assume that the opponent is the a reactive player $q = (q_1, q_2, q_1, q_2)$. By substituting $q_3 = q_1$ and $q_4 = q_2$ equations (22) and (23) are know re written as follow,

$$\begin{bmatrix} -q_2\left(4q_1 - 5q_2 - 1\right) \\ \left(q_2 - 1\right)\left(4q_1 - 5q_2 - 1\right) \end{bmatrix}$$

The sign of both equations is now based on the same term,$(4q_1 - 5q_2 - 1)$, which is a term that can have both negative and positive values. This is shown by Figure 5. Following this the following result is retrieved,

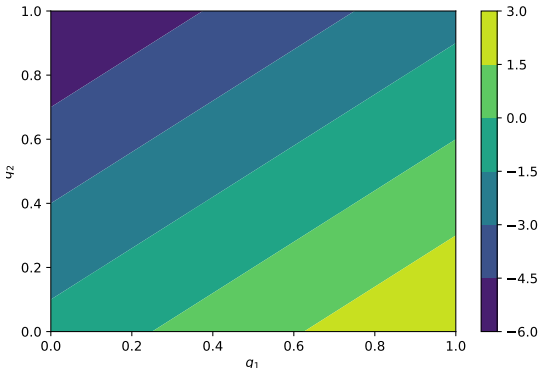


Figure 5: Sign of $(4q_1 - 5q_2 - 1)$.

Lemma 6 *Against a given reactive opponent $q = (q_1, q_2, q_1, q_2)$ defection is said to be stationary if and only if $(4q_1 - 5q_2 - 1)$ is negative.*

In tournaments the derivative 12 when we substitute for p_0 the equation is given by:

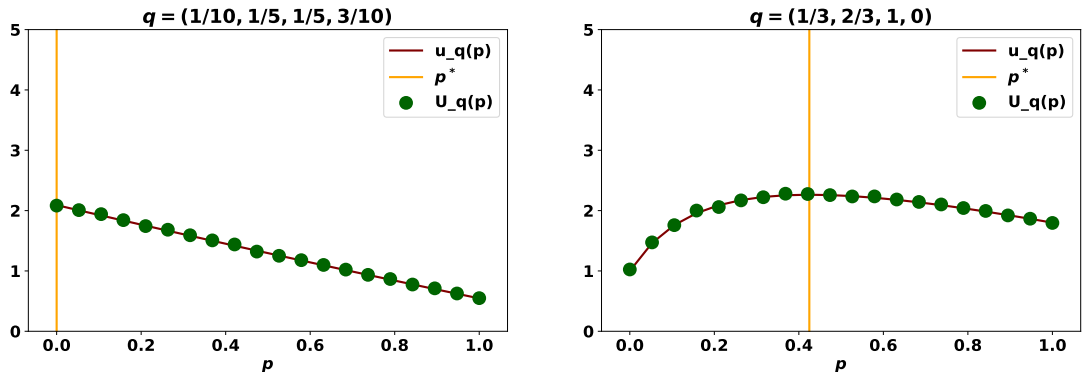


Figure 6: Numerical experiments for Algorithm 1 for $N = 1$.

$$\sum_{i=1}^{N}(c^{(i)T}\bar{a}^{(i)} - \bar{c}^{(i)T}a^{(i)}) \prod_{\substack{j=1 \\ j\neq i}}^{N} (\bar{a}^{(i)})^2 \tag{24}$$

The product is know to be positive thus in order for defection to be stable in a tournament falls down to a similar case the match one. Now the sum of columns must be negative.

4 Numerical Experiments

In this section several analytical results of Section 3 are validated using numerical experiments. Both methods and results are discussed. Initially, numerical methods for best responses in the constrained versions of the problem are presented. These further constrained problems taken into account in this work were:

- purely random strategies
- reactive strategies.

Moreover the answer to the questions discussed in Section 3.2 are answered via numerical algorithms. The question risen was identifying the best response memory one strategy.

4.1 Purely Random Strategies

Best responses of purely random strategies were given by Lemma 3 as presented in Section 3.3. Based on the results Algorithm 1 is constructed to perform a number of numerical experiments for pairwise and multi agent interactions.

The results of pairwise interactions are given by Figure 6. There is it evident that the optimal behaviour has been captured by our search algorithm. Moreover, the algorithm is also validated for tournament interactions, as shown by Figure 7.

Algorithm 1 Best response algorithm for purely random strategies

procedure PURELY RANDOM SEARCH
    $N \leftarrow$ number of opponets
    $S_q \leftarrow \{0, 1\}$
    $u' \leftarrow \frac{d \sum_{i=1}^{N} u}{d\bar{p}}$
    $\frac{u_N}{u_D} \leftarrow u'$
    $C(u_N) \leftarrow$ companion matrix of u_N
loop $i = 1$ to $2N$:
    $\lambda_i \leftarrow$ eigenvalue of $C(u_N)$
    if $u_D(\lambda_i) \neq 0$ **then**
        $S_q \cup \lambda_i$.
    goto *loop*.
    close;
    $p^* \leftarrow \text{argmax}(\sum_{i=1}^{N} u_{q^{(i)}}(p)), p \in S_q$.

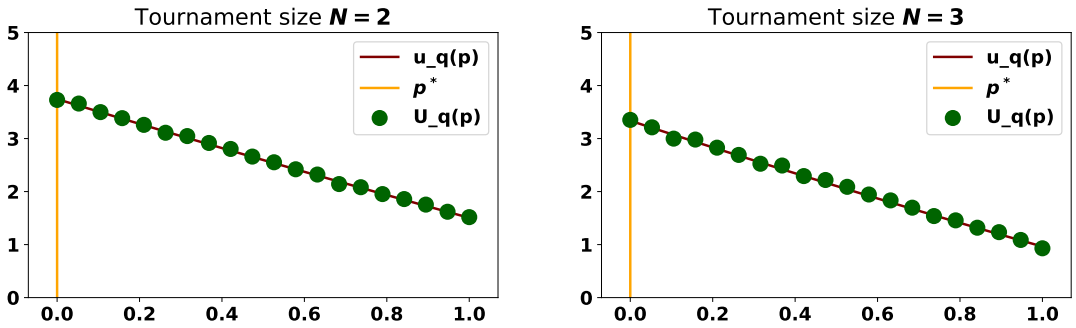


Figure 7: Numerical experiments for Algorithm 1 for $N > 1$.

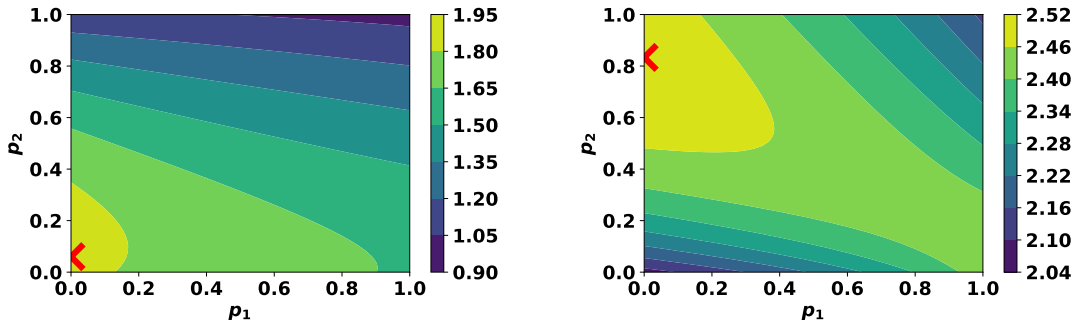


Figure 8: Numerical experiments for Algorithm 2 for $N = 1$.

4.2 Reactive Strategies

Best responses of reactive strategies have been discussed in Section 3.4. There it was stated that the field of resultant theory would be used to solve the partial derivatives of the utility. As reminder, best responses in reactive build upon Lemma 2 however now condition (13) is a 2 polynomial system of 2 variables.

Based on the results Algorithm 1 is constructed to perform a number of numerical experiments for pairwise and multi agent interactions.

Algorithm 2 Best response algorithm for reactive strategies

procedure REACTIVE SEARCH
  $N \leftarrow$ number of opponets
  $S_q \leftarrow \{0,1\}^2$
  $u' \leftarrow \frac{d \sum_{i=1}^{N} u}{d\bar{p}}$
  $\frac{u_N}{u_D} \leftarrow u'$
  $S(u_N, p_2) \leftarrow$ Sylvester's matrix forp_2.Coefficients are polynomials of p_1
  $R_S(S) \leftarrow det(S)$
  $\text{roots}_{p_1} \leftarrow p_1$ for $det(M)_{p_1} = 0$
loop root in roots_{p_1}:
   $\text{system}(p_2) \leftarrow u_N(root)$
  $\text{root}_{p_2} \cup p_2$ for $\text{system}(p_2) = 0$
  if $u_D((\text{root}, \text{root}_{p_2})) \neq 0$ **then**
    $S_q \cup \{(\text{root}, \text{root}_{p_2})\}$
  goto *loop*.
  close;
  $p^* \leftarrow \text{argmax}(\sum_{i=1}^{N} u_{q^{(i)}}(p)), p \in S_q$.

Illustrated by Figure 8 are the results of the numerical experiments for pairwise interactions. The results suggest that the best response behaviour is captured by our algorithm.

4.3 Memory one strategies

As discussed in Section the best response in memory one strategies will be captured using a numerical method. In this subsection the reasons for choosing our method as well as the parameters are presented.

Moreover, a parameter sweep has been performed and here we wil discuss the results. The algorithm used in order to maximise problem is bayesian optimisation.

Bayesian optimisation is ...

5 Limitation of memory

Best response behaviour of memory one strategies has been successfully captured as discuss in Section 4.3. The second part of this manuscript is focused on test the limitations of those very best responses.

This is achieved by comparing their performance with more complex strategies. In this section the methodology is covered as well as the results.

5.1 Methodology

In order to compare memory one strategies with complex strategies we perform the following steps. Tournaments of size $N = 3$ are considered, as those are the smallest possible tournaments. For a number of tournaments using the results of Section, we identify the best memory one strategy and it's utility for the given environment.

Afterwards, the memory on strategy is removed from the tournament and it is replaced with a more complex strategy. A more complex strategy is a strategy with a memory greater than one. The utility of that strategy is calculated adn then compared to that of the memory one strategy. The procedure is also described in Figure.

Though several more complex strategies could be used for comparison an 'archetype' was chosen here able to be trained. Several representation and training archetypes are described in [11]. We selected an archetype as it can be trained for the given environments similar to optimisation a memory one strategies for the given environments. From the archetypes described in [11] the chosen one called gambler and it is an archetypes that. Note that several different parameters have been used from different instances

of the gambler, so the strategies are almost generic. For training gambler the same method such as was used, the Bayesian optimisation.

The results of a big parameter sweep are discussed in the following section.

5.2 Limitation Results

Results, results, results.

6 Discussion

References

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

[2] Alkiviadis G. Akritas. *Sylvester's form of the Resultant and the Matrix-Triangularization Subresultant PRS Method*, pages 5–11. Springer New York, New York, NY, 1991.

[3] Howard Anton and Chris Rorres. *Elementary Linear Algebra: Applications Version*. Wiley, eleventh edition, 2014.

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

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

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

[7] Amir Beck and Marc Teboulle. A convex optimization approach for minimizing the ratio of indefinite quadratic functions over an ellipsoid. *Mathematical Programming*, 118(1):13–35, 2009.

[8] Hongyan Cai, Yanfei Wang, and Tao Yi. An approach for minimizing a quadratically constrained fractional quadratic problem with application to the communications over wireless channels. *Optimization Methods and Software*, 29(2):310–320, 2014.

[9] Alan Edelman and H Murakami. Polynomial roots from companion matrix eigenvalues. *Mathematics of Computation*, 64(210):763–776, 1995.

[10] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007.

[11] Marc Harper, Vincent Knight, Martin Jones, Georgios Koutsovoulos, Nikoleta E. Glynatsi, and Owen Campbell. Reinforcement learning produces dominant strategies for the iterated prisoners dilemma. *PLOS ONE*, 12(12):1–33, 12 2017.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] 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.

[18] 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.