

# 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:

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

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). Constraint (2) ensures that defection dominates cooperation and constraint (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  $(3, 1, 0, 5)$  [3].

$$T > R > P > S \quad (2)$$

$$2R > T + S \quad (3)$$

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 [4, 5] it attracted the attention of the scientific community.

In [4] a computer tournament of the IPD was performed. A tournament is a series of rounds of the PD between pairs of strategies. The topology commonly used, [4, 5], is that of a round robin where all contestants compete against each other. The winner of these tournaments 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. The robustness can also be checked through evolutionary process [14]. However, this aspect will not be considered here, instead the focus is on performance in tournaments.

In Axelrod’s original tournaments [4, 5], 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 use, 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 [12] and further used in [11, 13].

In [15] 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” [16]. In [16], 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 [10], 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

The purpose of this work is to consider a given memory one strategy in a similar fashion to [15]. However whilst [15] 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.

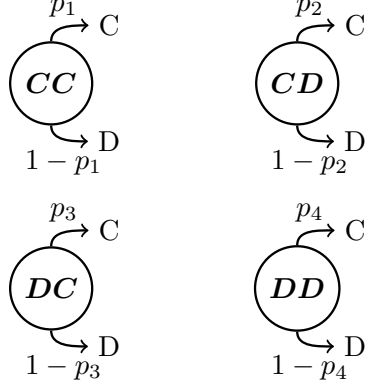
The second part of this manuscript we explore the limitation of the best response memory one strategies by comparing them to more complex strategies with a larger memory.

### 2.1 Background

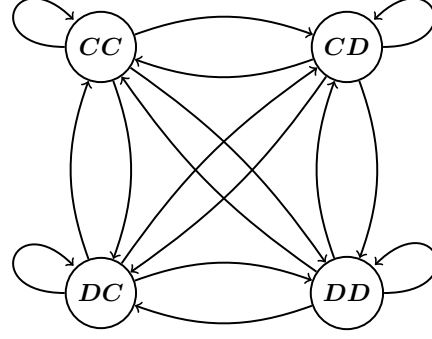
In this manuscript we explore the robustness of memory one strategies. A memory one strategy is defined as a strategy that decides it’s action in turn  $m$  based on what occurred in turn  $m - 1$ . If a strategy is concerned with only the outcome of a single turn then there are four possible ‘states’ 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,  $p = p_1, p_2, p_3, p_4 \in \mathbb{R}_{[0,1]}^4$ . A diagrammatic representation of such as strategy is given in Figure 1a.

In [13] a framework was introduced to study the interactions of memory one strategies modelled as a stochastic process, where the players move from one of the states  $CC, CD, DC, CC$  to another. More specifically, it can be modelled by the use of a Markov process of four states, shown by Figure 1b.

The transition matrix of the markov chain in Figure 1b is defined as  $M$  and is given by,



(a) Diagrammatic representation of a memory one strategy.



(b) Markov chain on a PD game.

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

Let the vector of the stationary probabilities of  $M$  be defined as  $v$ . 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,

$$u_q(p) = v \times S_p. \quad (5)$$

## 2.2 Utility

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.

Note though that  $u_q(p)$  is a function of 4 variables which is also affected by the transition probabilities of the opponent  $q$ . The first theoretical result that we introduce in this work is a compact way of writing  $u_q(p)$ . This is given by the Theorem 1.

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

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

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

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

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

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

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

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

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

## 2.3 Validation

In this section we validate the formulation of Theorem 1 using numerical experiments. All the simulated results of this work are done using [1] which is an open research framework for the study of the IPD. This package is described in [9].

To validate the formulation of  $u_q(p)$  several memory one players were matched against 20 opponents. 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

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 its utility  $u_q(p)$  against an opponent  $q$ . The decision variable is the vector  $p$  and the solitary constrains  $p \in \mathbb{R}_{[0,1]}^4$ . The optimisation problem is given by (11).

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

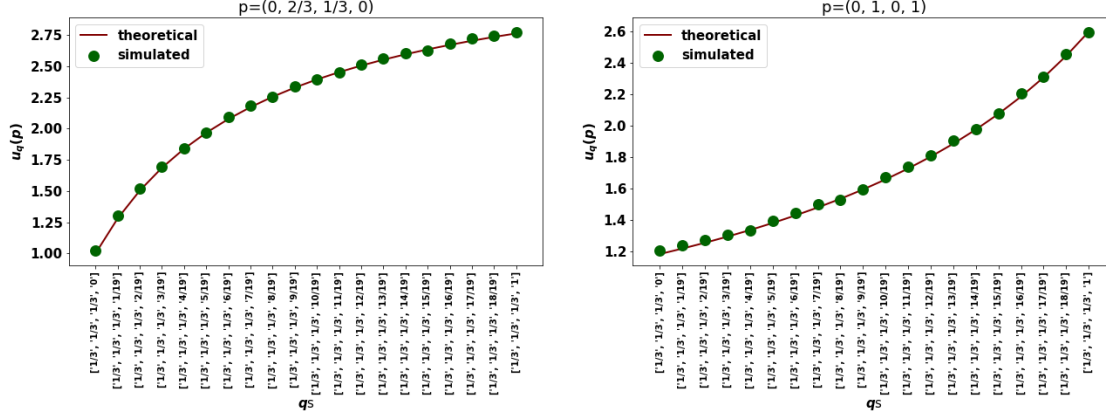


Figure 2: Differences between simulated and analytical results.

### 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 [8], will be 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. Optimising a non concave function is rather tricky.

Several articles in fractional optimisation of quadratic forms that was non concave can be found [6, 7]. Though in these works both the numerator and denominator of the fractional problem were concave. In [2] 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 Matches

The non concavity of  $u(p)$  indicates multiple local optimal points. Thus a compact way of searching the candidate optimal points needs to be introduced. Once the method is defined then the utility of each point is compared to the rest. The optimal point is the point that has the highest value of  $u(p)$ .

Our problem is a constrained optimisation problem and it is known that for such problems the set of candidate optimal solutions consists by edge cases and the roots of the utility's derivative. Edges case are all the possible combinations of  $p_1, p_2, p_3, p_4 \in \{0, 1\}$ .

Because of the non convexity we know we that there is a set of candidnate solutions. The candidnate solution with the maximum utility is also the optimal. Furthermore, we have a bounded problem, this mean that the we know that the candidnate solutions will be either on the bounds of our feasible solution in a point in the center. Note that the points are the roots of the derivate  $dudp$ .

The derivative of  $u(p)$  is calculated as follows,

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

Thus we can conclude that best response memory one strategy is given by Theorem.

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

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

where the set  $S_q$  is defined as

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

where the set  $\bar{p}$  is defined as the vector for which the following condition is true:

$$pQp\bar{Q}p^T - p\bar{Q}pQp^T + p(Q\bar{c}^T - \bar{Q}c^T)p + 2p(Q\bar{a} - \bar{Q}a + (c^T\bar{Q} - \bar{c}^TQ)p^T) = 2(\bar{c}^T a - c^T \bar{a})$$

where,

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

This can not be handle analytically further. The roots of the derivate are given by solving a multivariare system of 4 variavles where n=4. No further consideration was given into founding a propiate analytical method for that. This is mainle because the problem is already not as handaler that further more these are oble for the matches, where are in this work we want to consider tournament interactions are well.

The derivative of the a tournament is given by,

### 3.3 Tournament

Our approach can then be generalised to consider any given number of opponents. Against a set of opponents we want to optimise against the average utility, as it was proven by Lemma ?? that optimising against the average player is not representative. Thus the utility is given by,

$$\frac{1}{N} \sum_{i=1}^N u_q^{(i)}(p) = \frac{1}{N} \frac{\sum_{i=1}^N (\frac{1}{2}pQ^{(i)}p^T + c^{(i)T}p + a^{(i)}) \prod_{\substack{j=1 \\ j \neq i}}^N (\frac{1}{2}p\bar{Q}^{(j)}p^T + \bar{c}^{(j)T}p + \bar{a}^{(j)})}{\prod_{i=1}^N (\frac{1}{2}p\bar{Q}^{(i)}p^T + \bar{c}^{(i)T}p + \bar{a}^{(i)})}. \tag{13}$$

The derivative of which can be calculated as follow,

$$\begin{aligned} \frac{d}{dp} \frac{1}{N} \sum_{i=1}^N u_q^{(i)}(p) = \\ = \frac{(\sum_{i=1}^N Q_N^{(i)'} \prod_{\substack{j=1 \\ j \neq i}}^N Q_D^{(i)} + \sum_{i=1}^N Q_D^{(i)'} \sum_{\substack{j=1 \\ j \neq i}}^N Q_N^{(i)} \prod_{\substack{l=1 \\ l \neq i}}^N Q_D^{(i)}) \times \prod_{i=1}^N Q_D^{(i)} - (\sum_{i=1}^N Q_D^{(i)'} \prod_{\substack{j=1 \\ j \neq i}}^N Q_D^{(i)}) \times (\sum_{i=1}^N Q_N^{(i)} \prod_{\substack{j=1 \\ j \neq i}}^N Q_D^{(i)})}{(\prod_{i=1}^N Q_D^{(i)})^2} \end{aligned}$$

where,

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

More specifically the derivative of the smallest possible tournament of size 3 is given by:

It can be seen that the size of the problem is gradually increasing as opponents are being added to the tournament. Though no further consideration can be put into these in the following section we will introduce several numerical ways of getting the best response to these.

However this formulation still gives us insights. One of those is the stability of defection.

Also for constraint versions such as purely random player where  $p_1 = \dots$  the following two lemmas can be extracted. Both

In the following section we introduce a numerical method which will be used in following section.

## 4 Best responses Numerically

In this section we introduce several numerical methods. Numerical methods are used to supplement our analytical approach.

Initially we would like to present several insights and exact methods considered in this work to identify best response in constrained problems. These are in the purely random case and in the reactive case.

The numerical methods results.

## 5 Limitation of memory

## 6 Stability of defection

## References

- [1] The Axelrod project developers . Axelrod: [release title], April 2016.
- [2] Howard Anton and Chris Rorres. *Elementary Linear Algebra: Applications Version*. Wiley, eleventh edition, 2014.
- [3] R Axelrod and WD Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.
- [4] Robert Axelrod. Effective choice in the prisoner’s dilemma. *The Journal of Conflict Resolution*, 24(1):3–25, 1980.
- [5] Robert Axelrod. More effective choice in the prisoner’s dilemma. *The Journal of Conflict Resolution*, 24(3):379–403, 1980.
- [6] 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.
- [7] 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.
- [8] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Martin Nowak and Karl Sigmund. Game-dynamical aspects of the prisoner’s dilemma. *Applied Mathematics and Computation*, 30(3):191–213, 1989.
- [13] Martin Nowak and Karl Sigmund. The evolution of stochastic strategies in the prisoner’s dilemma. *Acta Applicandae Mathematicae*, 20(3):247–265, 1990.
- [14] Martin A Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge: Harvard University Press.
- [15] 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.
- [16] 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.