Good reactive strategies in the iterated prisoner's dilemma

Nikoleta E. Glynatsi, Christian Hilbe, Martin Nowak

We consider the infinitely repeated prisoner's dilemma with two players, denoted as p and q. At each turn, players p and q can choose between two actions; cooperate (C) or defect (D). A player who cooperates pays a cost $c > 0$ to provide a benefit $b > c$ for the co-player. A cooperator either gets $b - c$ (if the co-player also cooperates) or $-c$ (if the co-player defects). Respectively, a defector either gets b (if the co-player cooperates) or 0 (if the co-player defects). A strategy for player p is a mapping from the entire history of play to an action. There are infinitely many strategies that p can choose from, however, it is commonly assumed that the players' strategies depend on the last n turns only. These strategies are called memory-n strategies. A subset of memory-n strategies are $n-$bit reactive strategies. These are strategies that consider the last n action(s) of the co-player, opposed to considering their own actions as well.

Here we focus on two-bit reactive strategies ($n = 2$). There are 16 possible outcomes, which we denote as $E_pE_q|F_pF_q$ ($E_p, E_q, F_p, F_q \in \{C, D\}$) where the outcome of the previous round is E_pE_q and the outcome of the current round is F_pF_q. With the outcomes listed in order as $CC|CC, CC|CD, \dots, DD|DC, DD|DD$ a two-bit reactive strategy for p is a vector $\mathbf{p} = (p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4, p_1, p_2, p_1, p_2, p_3, p_4, p_3, p_4)$ where p_1 is the probability of cooperating when the last two actions of the co-player were C and C, p_2 is the probability of cooperating when the last two actions of the co-player were C and D, and so on. For simplicity, we denote a two-bit reactive strategy for p as $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$.

Let $\mathbf{v}$ be a probability distribution on the set of outcomes. $\mathbf{v}$ is a non-negative vector with unit sum indexed by the sixteen outcomes/states and it gives the probabilities that the players are in any of the states in the long run of the game. For example, v_3 is the probability that p played CD in the last two turns and q played CC. With respect to $\mathbf{v}$ we can define the expected payoffs for p and q, denoted as $\mathbf{s}_p$ and $\mathbf{s}_q$. We base the payoffs on the outcome of the last turn, and so $\mathbf{s}_p = \mathbf{v} \cdot \mathbf{S}_p$ and $\mathbf{s}_q = \mathbf{v} \cdot \mathbf{S}_q$ where,

$$\begin{aligned} \mathbf{S}_p &= (b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0, \ b-c, \ -c, \ b, \ 0) \quad \text{and} \\ \mathbf{S}_q &= (b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0, \ b-c, \ b, \ -c, \ 0). \end{aligned} \tag{1}$$

We derive a relationship between a player's two-bit reactive strategy and the resulting invariant distribution of the repeated game. This is given by Lemma 0.1.

Lemma 0.1. Assume that player p uses a two-bit reactive strategy $\hat{\mathbf{p}}$, and q uses a strategy that leads to a sequence of distributions $\{\mathbf{v}^{(n)}, n = 1, 2, ...\}$ with $\mathbf{v}^{(k)}$ representing the distribution over the states in the k^{th} round of the game. Let $\mathbf{v}$ be the associated stationary distribution, and let $\tilde{\mathbf{p}} = \hat{\mathbf{p}} - \hat{\mathbf{e}}_{12}$ where $\hat{\mathbf{e}}_{12} = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0)$. Then,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{v}^{(k)} \cdot \tilde{\mathbf{p}} = 0, \text{ and therefore } \mathbf{v} \cdot \tilde{\mathbf{p}} = 0.$$

That is,

$$\begin{aligned} (v_1 + v_9)(1 - \hat{p}_1) + (v_2 + v_{10})(1 - \hat{p}_2) + (v_5 + v_{13})(1 - \hat{p}_3) + (v_6 + v_{14})(1 - \hat{p}_4) \\ + (v_3 + v_{11})\hat{p}_1 + (v_4 + v_{12})\hat{p}_2 + (v_7 + v_{15})\hat{p}_3 + (v_8 + v_{16})\hat{p}_4 = 0. \end{aligned} \tag{2}$$

Proof. Akin in [Akin, 2016] derived a similar relationship when p played a memory-one strategy (Theorem 1.3). This is an extension of his work and it's proven completely analogously as in Akin's original paper. □

1 Good two-bit reactive strategies

The purpose here is to characterize the two-bit reactivate strategies that allow for Nash equilibria in which both players cooperate. Akin refers to these strategies as *good*, and here we use the same term. The definition is as follows.

Definition 1.1. Let a two-bit reactive strategy be **agreeable** if it is never the first to defect, and if it always cooperates with a probability 1 if the co-player has consecutively cooperated in that last two turns.

A strategy for p is called **good** if (i) it is agreeable, and (ii) if for any general strategy chosen by q against it the expected payoffs satisfy:

$$s_{\mathbf{q}} \geq b-c \qquad \Rightarrow \qquad s_{\mathbf{q}} = s_{\mathbf{p}} = b-c, \tag{3}$$

and the strategy is of **Nash type** if (i) it is agreeable and (ii) if for any general strategy chosen by q against it the expected payoffs satisfy:

$$s_{\mathbf{q}} \geq b-c \qquad \Rightarrow \qquad s_{\mathbf{q}} = b-c. \tag{4}$$

A good strategy is of Nash type, but not all strategies that are Nash are good.

Given the above, in the case of two-bit reactive strategies, we prove the following:

Theorem 1.1. Let the two-bit reactive strategy $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ be an **agreeable strategy**; that is $\hat{p}_1 = 1$. Strategy $\hat{\mathbf{p}}$ is **Nash** if the following inequalities hold:

$$\hat{p}_4 \leq 1 - \frac{c}{b} \qquad \hat{p}_2 \leq \hat{p}_4 \qquad \hat{p}_3 \leq 1 \qquad 1 + \hat{p}_2 \leq \frac{b}{c} - \hat{p}_4 \frac{b-c}{c}$$

The agreeable strategy $\hat{\mathbf{p}}$ is good if the inequalities above are strict.

Proof. We first eliminate the possibility $\hat{p}_4 = 1$. If $\hat{p}_4 = 1$, then $\hat{\mathbf{p}} = (1, \hat{p}_2, \hat{p}_3, 1)$. If q plays AllD $= (0,0,0,0)$ against $\hat{\mathbf{p}}$, then $v_6 = v_{CD|CD} = 1$. For $v_6 = 1$, $s_{\mathbf{q}} = b$ and $s_{\mathbf{p}} = -c$, and hence, $\hat{\mathbf{p}}$ is not of Nash type. We now assume $1 - \hat{p}_4 > 0$. Observe that,

$$\begin{aligned} s_{\mathbf{q}} - (b-c) &= \mathbf{v} \cdot \mathbf{S}_q - (b-c) \sum_{i=1}^{16} v_i \\ &= (v_2 + v_6 + v_{10} + v_{14})c + (c-b)(v_4 + v_8 + v_{12} + v_{16}) - b(v_3 + v_7 + v_{11} + v_{15}). \end{aligned} \tag{5}$$

Multiplying by the positive quantity $(1 - \hat{p}_4)$ and collecting terms, we have

$$\begin{aligned} &s_{\mathbf{q}} - (b-c) \geq 0 \Rightarrow \\ &(1-\hat{p}_4)(v_6 + v_{14})c \geq -c(1-\hat{p}_4)(v_2 + v_{10}) + (1-\hat{p}_4)(-c+b)(v_4 + v_8 + v_{12} + v_{16}) + b(1-\hat{p}_4)(v_3 + v_7 + v_{11} + v_{15}). \end{aligned} \tag{6}$$

Since $\tilde{p_1} = 0$, equation (2) implies

$$(1-\hat{p}_4)(v_{14}+v_6) = -((1-\hat{p}_2)(v_{10}+v_2)+(1-\hat{p}_3)(v_{13}+v_5)-\hat{p}_2(v_{12}+v_4)-\hat{p}_3(v_{15}+v_7)-\hat{p}_4(v_{16}+v_8)-v_{11}-v_3).$$

Substituting this in the above inequality and collecting terms we get,

$$A(v_{10}+v_2)+B(v_{12}+v_4)+C(v_{13}+v_5)+D(v_{15}+v_7)+E(v_{11}+v_{16}+v_3+v_8)\geq 0 \tag{7}$$

with

$$A=c(\hat{p}_2-\hat{p}_4), \qquad B=c(1+\hat{p}_2-\hat{p}_4)+b(-1+\hat{p}_4), \qquad C=c(-1+\hat{p}_3),$$
$$D=c\hat{p}_3+b(-1+\hat{p}_4), \qquad E=c+b(-1+\hat{p}_4).$$

For $A,B,C,D,E\leq 0$ we derive the following conditions,

$$\hat{p}_4\leq 1-\frac{c}{b} \qquad \hat{p}_2\leq\hat{p}_4 \qquad \hat{p}_3\leq 1 \qquad 1+\hat{p}_2\leq\frac{b}{c}-\hat{p}_4\cdot\frac{b-c}{c} \tag{8}$$

In the case where A,B,C,D and E are strictly smaller than 0, condition (7) holds iff $v_2,v_3,v_4,v_5,v_7,v_8,v_{10},v_{11},v_{12},v_{13},v_{15},v_{16}=0$. This implies, that $(v_1+v_9)(1-\hat{p}_1)+(v_6+v_{14})(1-\hat{p}_4)=0$. $\hat{p}_4$ can not be 1, thus $v_6,v_{14}=0$. This means $(v_1+v_9)=1$, so both players receive the reward payoff $(b-c)$ and $\hat{\mathbf{p}}$ is good. □

1.1 Open Question

A numerical exploration of Nash strategies in the space of two-bit reactive strategies made us conclude that the inequalities (8) are sufficient for a point to be Nash but not necessary. The numerical process and the results are summarised in Figure 1.

Note that since a two-bit reactive strategy $\hat{\mathbf{p}}=(\hat{p}_1,\hat{p}_2,\hat{p}_3,\hat{p}_4)$ can only be a Nash equilibrium if *no* other strategy yields a larger payoff, in particular neither AllD nor the Alternator strategy must yield a larger payoff, where AllD$=(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$ and Alternator$=(0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1)$. We conclude that an agreeable two-bit reactive strategy $\hat{\mathbf{p}}=(\hat{p}_1,\hat{p}_2,\hat{p}_3,\hat{p}_4)$ can only form a Nash equilibrium if

$$\pi(\text{AllD},\hat{\mathbf{p}})\leq b-c \quad\text{ and }\quad \pi(\text{Alternator},\hat{\mathbf{p}})\leq b-c,$$

or equivalently, if

$$\hat{p}_4\leq 1-\frac{c}{b} \quad\text{ and }\quad \hat{p}_2+\hat{p}_3\leq 1+\frac{b-c}{c} \tag{9}$$

In fact, the numerical analysis suggests the following stronger result.

Conjecture 1.2. An agreeable two-bit reactive strategy $\hat{\mathbf{p}}=(\hat{p}_1,\hat{p}_2,\hat{p}_3,\hat{p}_4)$ is of Nash type if and only if conditions (9) hold.

The question that remains is whether we can prove Conjecture 1.2.

References

E. Akin. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, pages 77–107, 2016.

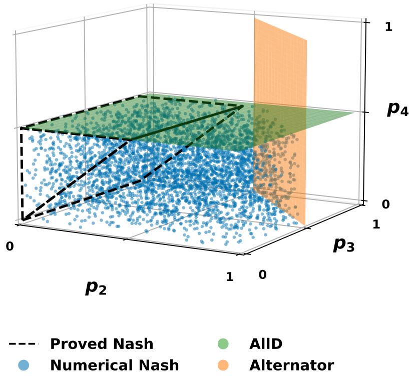

Figure 1: **Nash results for two-bit strategies.** The black dash line outlines the space that satisfies conditions (8). Thus, any two-bit reactive strategy that falls within this space is a good strategy. We have also numerically explored which agreeable strategies are Nash numerically. Namely, for a given agreeable two-bit strategy, and we checked if condition $\pi(\mathbf{q}, \hat{\mathbf{p}}) \leq (b-c)$ was satisfied against all pure memory-two strategies ($\mathbf{q} \in \{0,1\}^{16}$). We recorded if the strategy was Nash or not. We repeated this step for 10^4 random strategies. The results indicate that there are strategies that are Nash outside the proven space. Note however, that there are two planes that constrain the numerical Nash. There are the planes with the equations $\hat{p}_4 = 1 - \frac{c}{b}$ and $\hat{p}_3 = 1 + \frac{b-c}{c} - \hat{p}_2$. We obtain these by solving $\pi(\mathbf{q}, \hat{\mathbf{p}}) = (b-c)$ for $q \in \{\text{AllD and Alternator}\}$. Parameters: $c = 1, b = 2$.