Invincible Strategies of Iterated Prisoner's Dilemma

Extended Abstract

Shiheng Wang
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
swangbv@cse.ust.hk

Fangzhen Lin
HKUST-Xiaoi Robot Joint Laboratory
The Hong Kong University of Science and Technology
flin@cs.ust.hk

ABSTRACT

The Iterated Prisoner's Dilemma (IPD) is a well-known benchmark for studying rational agents' long term behaviour such as how cooperation can emerge among selfish and unrelated agents that need to co-exist over long term. Many well-known strategies have been studied, from the simple tit-for-tat (TFT) made famous by Axelrod after his influential tournaments to more involved ones like zero determinant and extortionate strategies studied recently by Press and Dyson. In this paper, we consider what we call invincible strategies. These are ones that will never lose against any other strategy in terms of average payoff in the limit. We provide a simple characterization of this class of strategies, and discuss its relationship with some other classes of strategies.

KEYWORDS

Evolution of cooperation; Repeated games; Memory-one strategies; Invincible strategies

ACM Reference Format:

Shiheng Wang and Fangzhen Lin. 2019. Invincible Strategies of Iterated Prisoner's Dilemma. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

The Iterated Prisoner's Dilemma is a classic benchmark used to study rational agents' long term behavior. It involves two agents playing repeatedly the Prisoner's Dilemma (PD). In the PD, each player can choose between Cooperate (C) and Defect (D). If both choose C, they receive a payoff of R (rewards); If both choose D, they receive a payoff of P (penalty); If one chooses C and the other D, the defector receives a payoff of T (temptation to defect) and the cooperator receives a payoff of S (sucker's payoff). The assumption is that $T > R > P > S$.

The profile (D, D) is the dominant Nash equilibrium of this game. But both players receive a higher payoff of R if they decide to cooperate, hence the dilemma. There is no controversy about what a rational agent should do when playing the PD. However, if the game is repeated indefinitely, it is not clear which if any strategy is the best. In fact, it is easy to see that there is no one best strategy against all other possible ones [2].

Researchers from diverse disciplines have used the IPD to study the emergence of cooperation among unrelated agents [2]. In 2012 Press and Dyson [14] dramatically changed people's understanding of this game by deriving what they called *zero determinant* (ZD) strategies. Among them, of particular interests are what they called *extortionate* strategies that can enforce an extortionate linear relation between the players' scores. We will show that extortionate strategies are invincible, in the sense that no strategies can have a higher average payoff when they play against them.

As it turned out, extortionate strategies are not the only ones that are invincible. In this paper, we formally define the class of invincible strategies, and show that they can be characterized by three simple conditions.

Invincible strategies are interesting for at least the following four reasons. Firstly, they have a very clear and intuitive definitions - never lose a match. Secondly, they are surprisingly simple to characterize - our main technical result in this paper is that they can be captured by three simple conditions. Thirdly, they are closely related with some other well-studied strategies such as extortionate strategies and Akin's good strategies [1]. Finally, as we will show by some experiments, they can play some important roles during the evolution of the emergence of cooperation.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

2 ITERATED PRISONER'S DILEMMA

The IPD is the repeated PD under the constraint that $T > R > P > S$ and $2R > T + S$ so that cooperation pays off in the long run. In a repeated game like the IPD, a player's strategy is a function from histories of interactions to actions. Often one restricts strategies to some specific forms, such as Turing machines [5, 10, 12], finite automata [3, 6, 15, 17], ones with limited memories [4, 7, 11], and other forms of bounded rationality (e.g. [13]).

For the IPD, Press and Dyson proved that the player with the shortest memory set the rule of the game [14], and one needs only consider *memory-one* mixed strategies. A memory-one (mixed) strategy decides with certain probability what action to do based on the outcome of the previous round. Thus it can be defined by the probabilities p_{CC}, p_{CD}, p_{DC}, and p_{DD} of playing C when the previous outcomes are CC, CD, DC, and DD, respectively. In the following, we write X's (player 1's) strategy $\mathbf{p}$ as a tuple in the order $\mathbf{p} = (p_1, p_2, p_3, p_4) = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$, and Y's strategy as a tuple in the following order: $\mathbf{q} = (q_1, q_2, q_3, q_4) = (q_{CC}, q_{DC}, q_{CD}, q_{DD})$. Notice that the orders for X and Y are the same when they are viewed from the player's *own* perspective.

A probability distribution $\mathbf{v}$ on the set of outcomes is a non-negative vector $\mathbf{v} = (v_1, v_2, v_3, v_4) = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$ with unit sum: $v_1 + v_2 + v_3 + v_4 = 1$. Given an initial distribution, the probability distribution after the r-th iteration is noted by $\mathbf{v}^r$.

An effective way to study memory-one strategies is to view their interactions as Markov chains. Our following presentation follows mostly after [14] and [1].

If X uses initial probability p_0 (for playing C) and strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$, Y uses initial probability q_0 and memory-one strategy $\mathbf{q} = (q_1, q_2, q_3, q_4)$, then the probability distribution of the first iteration is $\mathbf{v}^1 = (p_0q_0, p_0(1-q_0), (1-p_0)q_0, (1-p_0)(1-q_0))$ and the successive outcomes follow a Markov chain with transition matrix given by:

$$\mathbf{M} = \begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix}$$

Each entry of $\mathbf{M}$ represents the probability of transition between different states, and $\mathbf{M}\mathbf{v}^r = \mathbf{v}^{r+1}$. If $\mathbf{M}\mathbf{v} = \mathbf{v}$, then we say that $\mathbf{v}$ is stationary. Following [1], we call $\mathbf{M}$ convergent when there is a unique stationary distribution vector for M. Although the sequence of $\mathbf{v}^i (i = 1, 2, \ldots)$ may cycle through several states and thus not converge, the sequence of the Cesaro averages $\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}^i\}$ of the outcome distributions always converges to some stationary distribution $\mathbf{v}$. That is, if $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbf{v}^k = \mathbf{v}$ then $\mathbf{M}\mathbf{v} = \mathbf{v}$.

3 INVINCIBLE STRATEGIES

Definition 3.1 (Invincible Strategy). Player X plays a memory-one strategy $\mathbf{p}$ against player Y who plays memory-one strategy $\mathbf{q}$. Player X and Y get the average payoff of s_X and s_Y respectively. A memory-one strategy $\mathbf{p}$ is invincible if against any other memory-one strategy $\mathbf{q}$, for any initial distribution v_0, the players' payoffs satisfy $s_X \geq s_Y$.

LEMMA 3.2. *Assume* $\mathbf{v} = (v_1, v_2, v_3, v_4)$ *is the unique stationary distribution. Under* $T > S$, $s_X \geq s_Y$ *iff* $v_2 \leq v_3$.

THEOREM 3.3. *A strategy* $\mathbf{p}$ *is invincible iff*

$$p_2 + p_3 \leq 1, p_4 = 0, p_2 \neq 1 \tag{1}$$

Invincible strategies account for a large proportion of all strategies. Half of firm strategies ($p_4 = 0$) are invincible since the hyper plane $p_2 + p_3 \leq 1$ bisects the 3D cube $(p_1, p_2, p_3) \in [0, 1]$ when $p_4 = 0$. Well-known strategies that are invincible include *Tit-for-Tat* $(1, 0, 1, 0)$ which equalizes the payoff of both players, and *Always Defect* which never allow itself to be taken advantage of.

THEOREM 3.4. *All extortionate strategies are invincible strategies.*

Extortionate strategies are also zero determinate. However, not all zero determinate strategies are invincible. For example, $\mathbf{p} = (3/7, 0, 5/7, 2/7)$ is a zero determinant strategy that sets the co-player's score to a fixed value, but itself can receive a lower payoff.

While we have shown that all extortionate strategies are invincible, not all invincible strategies are extortionate. For example, $\mathbf{p}$=(0.5,0.2,0.7,0) is invincible, but this strategy is neither zero determinant nor extortionate.

The invincible strategies include extortionate ones. But they also include nice ones like the TFT, which is also a *good* one according to Akin [1]. While not all invincible strategies are Akin's good strategies, all *agreeable* ones are.

Definition 3.5 (Akin's Good Strategies [1]). X's strategy $\mathbf{p}$ is agreeable if $p_1 = 1$. It is good if it is agreeable and for any strategy chosen by Y, we have that

$$\text{if } S_Y \geq R \text{ then } S_Y = S_X = R.$$

Notice that $p_1 = 1$ means that this strategy always cooperate when the previous outcome is *CC*. Such strategies are also called nice strategies, which are never the first to defect.

THEOREM 3.6. *If a strategy is both agreeable and invincible, then it is good.*

However, not all good strategies are invincible. This follows from Akin's Theorem 1.5 in [1].

4 EXPERIMENT

Since Press and Dyson's work, there have been several experiments about ZD and extortionate strategies. One was by Stewart and Plotkin [16] who ran a Axelrod-style tournaments that include a few extortionate strategies. One of the notable results of Stewart and Plotkin's tournament is that the extortionate strategy named Extort-2 won the second most head-to-head matches. We now know this is not really surprising given that extortionate strategies are invincible. Actually, no invincible strategy will loose a head-to-head match if the game is repeated for sufficient number of rounds.

Another was by Hilbe *et al.* [8] who ran an experiment to analyze the evolutionary performance of extortionate strategies. They concluded that extortionate strategies can act as catalysts for the evolution of cooperation but that they themselves are not the stable outcome of natural selection. We rerun their experiment and replace extortionate strategy with the invincible strategy (0.9,0.7,0.2,0), and the result turns out to be similar. Notice that (0.9,0.7,0.2,0) is just invincible. It is not even a zero determinant strategy.

The Axelrod python library [9] make it easy to run experiments about the IPD. We run four experiments similar to some of those in [8] but with extortionate strategies replaced by invincible ones.

The evolutionary behavior of invincible strategies turn out to be similar to that of extortionate ones.

(a) The population begins with cooperative strategy *Win-Stay-Lose-Shift* and defective strategy *Always Defect.* After 70 iterations, defective strategies dominates this population.

(b) After adding agents of invincible strategy (0.9,0.7,0.2,0), defectors are firstly eliminated, followed by invincible strategy. Cooperative strategy becomes the stable outcome. Thus, invincible strategy acts as catalyst of cooperation, but is not a stable outcome of evolution.

(c) In another experiment, several strategies evolve in one population, where all agents play the IPD with every other agents, cooperative strategies becomes the winner.

(d) After invincible strategies ally and play against the alliance of other strategies (agents in the same alliance won't play with each other), the population of invincible strategies dominates the other.

5 CONCLUSION

Inspired by our initial observation that no strategies can defeat an extortionate strategy, we went on to study the class of all such strategies that will never lose a head-to-head match and call them invincible. Our main technical result is a precise characterization of this class of strategies by three simple conditions. We have also related this class of strategies to others and considered their role in the emergence of cooperation. We have extended this class to a more general form of repeated games and will report our results elsewhere.

REFERENCES

[1] Ethan Akin. 2016. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems* (2016), 77–107.

[2] Robert Axelrod and William Donald Hamilton. 1981. The evolution of cooperation. *science* 211, 4489 (1981), 1390–1396.

[3] Elchanan Ben-Porath. 1990. The complexity of computing a best response automaton in repeated games with mixed strategies. *Games and Economic Behavior* 2, 1 (1990), 1–12.

[4] Lijie Chen, Fangzhen Lin, Pingzhong Tang, Kangning Wang, Ruosong Wang, and Shiheng Wang. 2017. K-Memory Strategies in Repeated Games. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1493–1498.

[5] Lijie Chen and Pingzhong Tang. 2015. Bounded rationality of restricted turing machines. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1673–1674.

[6] Itzhak Gilboa. 1988. The complexity of computing best-response automata in repeated games. *Journal of economic theory* 45, 2 (1988), 342–352.

[7] CH Hauert and Heinz Georg Schuster. 1997. Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: a numerical approach. *Proceedings of the Royal Society of London B: Biological Sciences* 264, 1381 (1997), 513–519.

[8] Christian Hilbe, Martin A Nowak, and Karl Sigmund. 2013. Evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences* 110, 17 (2013), 6913–6918.

[9] Vincent Knight, Owen Campbell, Marc Harper, Karol Langner, James Campbell, Thomas Campbell, Alex Carney, Martin Chorley, Cameron Davidson-Pilon, Kristian Glass, et al. 2016. An open reproducible framework for the study of the iterated prisoner's dilemma. *arXiv preprint arXiv:1604.00896* (2016).

[10] Vicki Knoblauch. 1994. Computable Strategies for Repeated Prisoner's Dilemma. *Games and Economic Behavior* 7, 3 (1994), 381–389.

[11] Kristian Lindgren. 1992. Evolutionary phenomena in simple dynamics. In *Artificial life II*. 295–312.

[12] Nimrod Megiddo and Avi Wigderson. 1986. On play by means of computing machines: preliminary version. In *Proceedings of the 1986 Conference on Theoretical aspects of reasoning about knowledge*. Morgan Kaufmann Publishers Inc., 259–274.

[13] Martin J Osborne and Ariel Rubinstein. 1994. *A course in game theory*. MIT press.

[14] William H Press and Freeman J Dyson. 2012. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* 109, 26 (2012), 10409–10413.

[15] Ariel Rubinstein. 1986. Finite automata play the repeated prisoner's dilemma. *Journal of economic theory* 39, 1 (1986), 83–96.

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

[17] Song Zuo and Pingzhong Tang. 2015. Optimal Machine Strategies to Commit to in Two-Person Repeated Games.. In *AAAI*. 1071–1078.