ARTICLE IN PRESS

Journal of Theoretical Biology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The robustness of zero-determinant strategies in Iterated Prisoner's Dilemma games

Q1 Jing Chen [a,*], Aleksey Zinger [b]

[a] Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA
[b] Department of Mathematics, Stony Brook University, Stony Brook, NY 11794, USA

HIGHLIGHTS

- We prove a strong theoretical result on the power of Zero-Determinant strategies.
- Here a strategic player and an adapting player play Iterated Prisoner's Dilemma.
- By using ZD strategies, the former guarantees the best score in all adapting paths.
- All adapting paths end up as if the latter unconditionally cooperates.
- The former is safe to use ZD strategies, even he does not know the other's reaction.

ARTICLE INFO

Article history:
Received 22 September 2013
Received in revised form
20 April 2014
Accepted 1 May 2014

Keywords:
Q3 Adapting paths

ABSTRACT

Press and Dyson (2012) discovered a special set of strategies in two-player Iterated Prisoner's Dilemma games, the *zero-determinant* (*ZD*) *strategies*. Surprisingly, a player using such strategies can unilaterally enforce a linear relation between the payoffs of the two players. In particular, with a subclass of such strategies, the *extortionate strategies*, the former player obtains an advantageous share of the total payoff of the players, and the other player's best response is to always cooperate, by doing which he maximizes the payoff of the extortioner as well. When an extortionate player faces a player who is not aware of the theory of ZD strategies and improves his own payoff by adaptively changing his strategy following some unknown dynamics, Press and Dyson conjecture that there always exist adapting paths for the latter leading to the maximum possible scores for both players.

In this work we confirm their conjecture in a very strong sense, not just for extortionate strategies, but for *all* ZD strategies that impose positive correlations between the players' payoffs. We show that not only the conjectured adapting paths always exist, but that actually *every* adapting path leads to the maximum possible scores, although some paths may not lead to the unconditional cooperation by the adapting player. This is true even in the rare cases where the setup of Press and Dyson is not directly applicable. Our result shows that ZD strategies are even more powerful than as pointed out by their discoverers. Given our result, the player using ZD strategies is assured that she will receive the maximum payoff attainable under the desired payoff relation she imposes, without knowing how the other player will evolve. This makes the use of ZD strategies even more desirable for sentient players.

1. Introduction

The two-player Iterated Prisoner's Dilemma (IPD) game is one of the standard models for studying the emergence of cooperative behavior among competitive players. It has long been investigated in economics, political science, evolutionary biology, and computer science (see Dawkins, 1976; Axelrod and Hamilton, 1981; Axelrod, 1984; Roberts, 1985; Axelrod and Dion, 1988; Rubinstein, 1986; Poundstone, 1992; Nowak and Sigmund, 1993; Papadimitriou and Yannakakis, 1994; Bhaskar and Obara, 2002; Ely and Välimäki, 2002; Nowak, 2006; Kendall et al., 2007; Bhaskar et al., 2008, as just a few examples). As IPD has been so widely studied, it was Q4 surprising when Press and Dyson (2012) discovered a completely new property of this game, namely, the existence of Zero-Determinant (ZD) strategies. Roughly speaking, such strategies allow one player to unilaterally set the payoff score of the other or to enforce a linear relation between the two players' scores, as opposite to the previous general belief that no ultimatum strategy can enforce any specific kind of outcome. Among such strategies, of

* Corresponding author. Tel.: +1 631 632 1827.
E-mail addresses: jingchen@cs.stonybrook.edu (J. Chen), azinger@math.sunysb.edu (A. Zinger).

http://dx.doi.org/10.1016/j.jtbi.2014.05.004
0022-5193/© 2014 Published by Elsevier Ltd.

		Player Y C	Player Y D
Player X	C	(R,R)	(S,T)
Player X	D	(T,S)	(P,P)

Fig. 1. Scores for the players X and Y in each of the four outcomes in a single play of Prisoner's Dilemma.

particular interest are the so-called extortionate strategies (Press and Dyson, 2012), in which the sentient player takes a larger share of the total benefit, and generous strategies (Stewart and Plotkin, 2012, 2013; Akin, 2013), in which the sentient player takes a larger share of the total loss from the full-cooperation rewards. The results of Press and Dyson (2012) have led to completely new viewpoints on IPD. Since then, the properties of ZD strategies, including extortionate strategies in arbitrary IPD games and generous strategies in donation games, have been actively studied; see Stewart and Plotkin (2012), Hilbe et al. (2013), Stewart and Plotkin (2013), Adami and Hintze (2013), Akin (2013), Szolnoki and Perc (2014), and Hilbe et al. (2013).

The game under consideration here is of discrete time and with infinitely many rounds. In each round, the same two players, X and Y, play the one-shot Prisoner's Dilemma (PD). As illustrated in Fig. 1, each player can choose to cooperate (C) or to defect (D), without knowing the other's choice. If both cooperate, then each receives score R. If both defect, then each receives a smaller score P. If one cooperates and the other defects, then the defector rips off a score T larger than R, and the cooperator gets ripped off with a score of S smaller than P. The literature typically assumes $2R > T+S > 2P$, so that the total score of the players is maximized when both cooperate. For example, $(T,R,P,S) = (5,3,1,0)$ is a conventional realization of the parameters.

Press and Dyson (2012) assume that both players have memory of length 1, i.e. what a player does in the current round only depends on the outcome of the previous round, rather than the whole history of the play or the number of rounds played. Accordingly, a (mixed) strategy of a player consists of a mapping from the four possible outcomes of PD to the probabilities of cooperating. The strategies of the two players together with a starting outcome determine a Markov chain. The players' payoff scores, s_X and s_Y, are defined to be the expected scores they would receive under the stationary distribution of the Markov chain.

A ZD strategy of player X guarantees

$$s_X - K = \chi(s_Y - K)$$

for specified values of χ and K satisfying certain conditions, no matter what strategy Y uses. For s_X and s_Y to be positively correlated, one needs $\chi \geq 1$ (the cases $\chi \in (0,1)$ do not correspond to ZD strategies). Facing such ZD strategies, when Y adjusts his own strategy to increase his score, he increases X's score even more, and when he achieves his own maximum score, X's score is also maximized. Both extortionate and generous strategies are *positively correlated ZD* (*pcZD*) *strategies*, with the former satisfying $K = P$ and the latter $K = R$.

As pointed out by Press and Dyson (2012), for extortionate strategies, the scores of both players are maximized when Y cooperates unconditionally, namely, uses the strategy $(1,1,1,1)$.

An extortionate player facing an adapting player. One question which is not completely answered by Press and Dyson (2012) and not considered by previous followups is the following. *What should a player X witting of extortionate strategies do if she believes that her opponent Y is an adapting player*? An adapting player is one who tries to improve his own score following some optimization scheme (perhaps known only to him), but without explicitly considering or trying to alter the strategy of X. Such a player is called an evolutionary player in Press and Dyson (2012), but "evolutionary" already has a specific (and different) meaning in the context of game theory, and thus we use "adapting" instead, to avoid confusion. The answer to the question above depends on how Y adapts. Although it is of Y's best interest to unconditionally cooperate, he may not realize this fact and may only make local movement to gradually improve his score. Since the direction of improvement is not unique, in principle Y might end up at a local optimum and leave X with a score much smaller than what she expects when Y unconditionally cooperates. If this can happen, then X would use an extortionate strategy only if she believes that Y will take a desirable adapting path (roughly speaking, an *adapting path* is a smooth map from time to Y's strategies such that Y's utility increases along time—formally defined in Section 3), and would otherwise continue monitoring the behavior of Y and change her strategy when necessary.

Press and Dyson conjecture that in all cases, that is, with different parameters (R,T,S,P), different starting points of IPD, and different original strategies of Y, there *exist* adapting paths of Y that lead to the globally maximum scores when X applies an extortionate strategy. However, the existence of desirable adapting paths is not sufficient for one to conclude that X should extort Y. If there are other adapting paths where Y ends up at a local optimum, it is unclear what X should do, as discussed above. In the numerical experiment of Press and Dyson (2012) for the conventional parameters mentioned before, the adapting paths examined do not end up at a local optimum, but formal analysis of the general case is missing.

The same question can be asked for all pcZD strategies, not only the extortionate ones.

Our contribution: We prove the conjecture of Press and Dyson (2012) in a very strong and general form, and analytically justify the use of extortionate as well as other pcZD strategies against adapting players. We show that in all cases, *all* adapting paths of Y lead to the maximum scores, although the strategy of Y may not end up at the unconditional cooperation. This holds even in some degenerate cases where the analysis of Press and Dyson (2012) does not apply. Accordingly, as long as Y does not stop at a locally sub-optimal strategy and does not evolve at a speed that goes to 0 as time goes to infinity, the dynamics will always end up at the maximum scores attainable under the linear relation imposed by X. Therefore, it is always "safe" for X to use pcZD strategies, and she will receive her desired score in a very robust way, without knowing which adapting path Y will follow.

As an easy consequence of our main result, if X does not want to take any advantage over Y, but instead is benevolent and wishes to promote mutual cooperation, she is able to do so in all cases, via a "fair" extortionate strategy, where $\chi = 1$, or via a generous strategy. In this way, X enforces the maximum total score of the two players, (R,R), which de facto is equivalent to the unconditional cooperation by both players. This is true even when Y only evolves selfishly and does not care about the total score at all.

Related work: The original setup of Press and Dyson (2012) is very different from that of all other studies of ZD strategies so far (Adami and Hintze, 2013; Akin, 2013; Hilbe et al., 2013; Stewart and Plotkin, 2013; Hilbe et al., 2013; Szolnoki and Perc, 2014). In particular, in Press and Dyson (2012) there are only two players, one of them uses a fixed ZD strategy and the other changes his strategy over time. While in all other studies, which focus on evolutionary aspects of ZD strategies, there are one or two

Please cite this article as: Chen, J., Zinger, A., The robustness of zero-determinant strategies in Iterated Prisoner's Dilemma games. J. Theor. Biol. (2014), http://dx.doi.org/10.1016/j.jtbi.2014.05.004