Zero-Determinant Strategies in Repeated Prisoner’s Dilemma Games

Genki Ichinose\textsuperscript{1} and Naoki Masuda\textsuperscript{2}

\textsuperscript{1}Department of Mathematical and Systems Engineering, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, 432-8561, Japan
\textsuperscript{2}Department of Engineering Mathematics, University of Bristol, Merchant Venturers Building, Woodland Road, Clifton, Bristol BS8 1UB, United Kingdom
ichinose@shizuoka.ac.jp

Abstract

Direct reciprocity is one of the mechanisms for sustaining mutual cooperation in repeated social dilemma games. Zero-determinant (ZD) strategies allow a player to unilaterally set a linear relationship between the player’s own payoff and the co-player’s payoff regardless of the strategy of the co-player. The original ZD strategies were derived for infinitely repeated games. Here, we analytically search for ZD strategies in finitely repeated prisoner’s dilemma games. Our results can be summarized as follows. First, we present the forms of ZD in finitely repeated games, which are directly extended from the known results for infinitely repeated games. Second, for the three most notable ZD strategies, the equalizers, extortioners, and generous strategies, we derive the threshold discount factor value above which the ZD strategies exist. Finally, we show that the only strategy sets that enforce a linear payoff relationship are either the ZD strategies or unconditional strategies.

The prisoner’s dilemma game is widely used for studying the emergence of cooperation where two individuals are involved in a social dilemma and each individual selects either cooperation (C) or defection (D). An individual obtains a larger payoff by selecting D no matter what the other individual does. However, mutual defection, which is the unique Nash equilibrium of the game, yields a smaller benefit to both players compared to mutual cooperation. Cooperation can be sustained in the prisoner’s dilemma game and other social dilemma games through various mechanisms (Nowak, 2006; Sigmund, 2010).

In 2012, Press and Dyson found a novel class of strategies in the repeated prisoner’s dilemma game, called zero-determinant (ZD) strategies (Press and Dyson, 2012). ZD strategies impose a linear relationship between the payoffs of a focal individual and its co-player regardless of the strategy that the co-player implements. The advent of the ZD strategies has opened new lines of investigation of direct reciprocity. Most of the previous studies for the ZD strategies have been conducted under the assumption of infinitely repeated games. Finitely repeated games are probably more realistic than infinitely repeated games and are compatible with experimental studies. In this presentation, we study the ZD strategies in the finitely repeated prisoner’s dilemma game. There are a few related studies which have investigated ZD strategies in finitely repeated games (Hilbe et al., 2015; McAvoy and Hauert, 2016; McAvoy and Hauert, 2017).

Here, we consider the finitely repeated prisoner’s dilemma game with memory-one strategies. The symmetric two-person prisoner’s dilemma game has a payoff matrix of

\[
\begin{pmatrix}
C & D \\
D & R & S \\
C & T & P
\end{pmatrix},
\]

where \( T > R > P > S \) and \( 2R > T + S \). The entries represent the payoffs that the focal player, denoted by \( X \), obtains in a single round of a repeated game. Each row and column corresponds to the action of the focal player, \( X \), and the co-player (denoted by \( Y \)), respectively. The two players repeat the game where subsequent rounds, given the current round, take place with probability \( w \) \((0 < w < 1)\), which is called the discount factor.

Consider two players \( X \) and \( Y \) that adopt memory-one strategies, with which they rely on only the outcome of the last round to decide the action to be submitted in the current round. A memory-one strategy is composed of a 5-tuple; \( X \)’s strategy is given by a combination of \( p = (p_{CC}, p_{CD}, p_{DC}, p_{DD}) \) and \( p_0 \), where \( 0 \leq p_{CC}, p_{CD}, p_{DC}, p_{DD}, p_0 \leq 1 \). \( p_{CC} \) is the conditional probability that \( X \) cooperates when both \( X \) and \( Y \) cooperated in the last round, \( p_{CD} \) is the conditional probability that \( X \) cooperates when \( X \) cooperated and \( Y \) defected in the last round, and so forth. \( p_0 \) denotes the probability that \( X \) cooperates in the first round. Similarly, \( Y \)’s strategy is composed of a combination of \( q = (q_{CC}, q_{CD}, q_{DC}, q_{DD}) \) and the probability to cooperate in the first round, \( q_0 \), where \( 0 \leq q_{CC}, q_{CD}, q_{DC}, q_{DD}, q_0 \leq 1 \). Because both players use a memory-one strategy, the stochastic state of the two players in round \( t \) \((t \geq 0)\) is described by \( \nu(t) = (\nu_{CC}(t), \nu_{CD}(t), \nu_{DC}(t), \nu_{DD}(t)) \), where \( \nu_{CC}(t) \) is the probability that both players cooperate in round \( t \), \( \nu_{CD} \) is the probability that \( X \) cooperates and \( Y \) defects,
and so forth. The expected per-round payoff to player \( X \) in the repeated game is calculated by \( \pi_X = (1 - w) \sum_{t=0}^{\infty} w^t v(t)S_X^t \) where \( S_X = (R, S, T, P) \). Similarly, the expected per-round payoff to player \( Y \) is calculated by \( \pi_Y = (1 - w) \sum_{t=0}^{\infty} w^t v(t)S_Y^t \) where \( S_Y = (R, T, S, P) \).

We search for player \( X \)'s strategies which impose a linear payoff relationship between the two players, i.e.,

\[
\alpha \pi_X + \beta \pi_Y + \gamma = 0. \tag{2}
\]

When \( \alpha \neq 0 \), we rewrite \( \chi = -\beta/\alpha \) and \( \kappa = -\gamma/(\alpha + \beta) \) to transform Eq. (2) to

\[
\pi_X - \kappa = \chi(\pi_Y - \kappa). \tag{3}
\]

By definition, the equalizer, which is one of the special cases of the ZD strategies, unilaterally sets the co-player’s payoff, \( \pi_Y \), to a constant value irrespective of the co-player’s strategy. To derive a solution for the equalizer strategies in the finitely repeated game, we proceed with the following idea: If a strategy \( p \) ensures that the payoffs of the two players are on a horizontal line in the \( \pi_X-\pi_Y \) space, regardless of the co-player’s strategy, then the payoffs must be on that horizontal line if the co-player uses unconditional cooperation or unconditional defection. Substituting the co-player’s unconditional cooperation and unconditional defection into the payoff equations provides the necessary conditions imposed on \( X \)'s strategy. A straightforward computation then shows that these necessary conditions are in fact also sufficient. Specifically, \( \pi_{Y;0000} \) is the payoff of \( Y \) when \( q = (0, 0, 0, 0) \). Similarly, \( \pi_{Y;1111} \) is the payoff of \( Y \) when \( q = (1, 1, 1, 1) \). If \( X \) adopts an equalizer strategy, \( \pi_{Y;0000} = \pi_{Y;1111} \) must hold true regardless of \( \eta_0 \). From this constraint, we derived the necessary conditions that the strategy is the equalizer strategy in the finitely repeated games. We inversely showed that those necessary conditions were also sufficient. The resulting expression of the equalizer in finitely repeated games is given by

\[
p_{CD} = p_{CC}(T - P) - \left( \frac{1}{w} + p_{DD} \right)(T - R), \tag{4}
\]

\[
p_{DC} = \left( \frac{1}{w} - p_{CC} \right)(P - S) + p_{DD}(R - S). \tag{5}
\]

These equations extend the results previously obtained for \( w = 1 \) (Press and Dyson, 2012).

We next identified the condition for \( w \) under which equalizer strategies exist. Owing to the constraints of \( 0 \leq p_{CC}, p_{DD} \leq 1 \), the condition is given by

\[
w \geq w_c \equiv \max \left( \frac{T - R}{T - P}, \frac{P - S}{P - R} \right). \tag{6}
\]

When the game is repeated sufficiently many times as specified in Eq. (6), the equalizer strategies exist. By further analyses, we showed that the condition was the same for the extortioner and the generous strategy.

All strategies except for the equalizer are given in the form of Eq. (3). Next, we derive the expressions of \( X \)'s strategy that realizes the equation. By applying the same idea used for the equalizer, we derived either

\[
w p = \begin{pmatrix}
1 - \phi(\chi - 1)(R - \kappa) - (1 - w)p_0 \\
1 + \phi(\chi(\chi - 1) - \kappa - \chi T + S) - (1 - w)p_0 \\
\phi(\chi - 1)(\kappa - P) - (1 - w)p_0
\end{pmatrix}, \tag{7}
\]

or

\[
p_0 = p_{CC} = p_{CD} = p_{DC} = p_{DD} \quad (0 \leq p_0 \leq 1) \tag{8}
\]

to satisfy Eq. (3). The former corresponds to the ZD strategies and the latter corresponds to the unconditional strategies in finitely repeated games, respectively. In the former case, \( \kappa = P \) is equal to the extortioner while \( \kappa = R \) is the generous strategy, respectively.

In conclusion, we analyzed ZD strategies in finitely repeated prisoner’s dilemma games with general payoff matrices. The prominent results derived in the present paper are two-fold. First, we derived the threshold discount factor value, \( w_c \), above which the ZD strategies exist for the equalizer, extortioner, and generous strategies. All of them share the same threshold value. Second, we revealed that the memory-one strategies that impose a linear relationship between the payoff of the two players are either ZD strategies or an unconditional strategy. The latter class includes the unconditional cooperator and unconditional defector as special sets. For further details, see our published paper (Ichinose and Masuda, 2018).

References


