Zero-Determinant Strategies under Observation Errors

Azumi Mamiya[1] and Genki Ichinose[1*]

[1]Department of Mathematical and Systems Engineering, Shizuoka University,
3-5-1 Johoku, Naka-ku, Hamamatsu, 432-8561, Japan
[*]ichinose.genki@shizuoka.ac.jp

Abstract

In 2012, Press and Dyson discovered a strategy set, called Zero-determinant (ZD) strategies, which enforces a linear payoff relationship between a focal player and the opponent regardless of the opponent's strategy in the repeated prisoner's dilemma (RPD) game. In the RPD game, a discount factor and observation errors are both important because they often happen in society. Here, we examined strategies that enforce linear payoff relationships in the RPD game considering both a discount factor and observation errors. As a result, we first revealed that the payoffs of two players can be represented by the form of determinants even with these two factors. Then, we searched for all possible strategies that enforce linear payoff relationships and found that both ZD strategies and unconditional strategies are the only strategy sets which satisfy the condition.

The prisoner's dilemma game is a model for studying the emergence of cooperation among competitive players. In a one-shot interaction, cooperation is not likely to occur because it costs the actor while defection does not. However, when the game is repeated, cooperation will be rewarded by the opponent in the future. In such a situation, cooperation becomes a possible equilibrium. Although the rules of the game are easy, predicting results is complex in the repeated prisoner's dilemma (RPD) game. In that context, Press and Dyson found a novel class of strategies called zero-determinant (ZD) strategies (Press and Dyson, 2012). ZD strategies impose a linear relationship between the payoffs of a focal player and its co-player regardless of the strategy that the co-player implements.

In the RPD game, considering a discount factor and observation errors is important because they often happen in society. Recent studies about ZD strategies have only focused on either one: a discount factor (Hilbe et al., 2015; Ichinose and Masuda, 2018) or observation errors (Hao et al., 2015; Mamiya and Ichinose, 2019). As far as we know, there is no study which considers both a discount factor and observation errors in ZD strategies. Therefore, we study ZD strategies in the situation that incorporates both a discount factor and observation errors.

We consider the symmetric two-person RPD game with private monitoring. Each player $i \in \{X, Y\}$ chooses an action $a_i \in \{\mathrm{C}, \mathrm{D}\}$ in each round, where C and D stand for cooperation and defection, respectively. After the two players executed the action, player i observes his own action a_i and private signal $\omega_i \in \{g, b\}$ about the opponent's action, where g and b stand for *good* and *bad*, respectively. In private monitoring, players sometimes misunderstand the signals. $\sigma(\boldsymbol{\omega}|\boldsymbol{a})$ is the probability that a signal profile $\boldsymbol{\omega} = (\omega_X, \omega_Y)$ is realized when the action profile is $\boldsymbol{a} = (a_X, a_Y)$. Let ϵ be the probability that an error occurs to one particular player but not to the other player while ξ be the probability that an error occurs to both players. Then, the probability that an error occurs to neither player is $1 - 2\epsilon - \xi$. In each round, player i's realized payoff $u_i(a_i, \omega_i)$ is determined by his own action a_i and signal ω_i, such that $u_i(\mathrm{C}, g) = R$, $u_i(\mathrm{C}, b) = S$, $u_i(\mathrm{D}, g) = T$, and $u_i(\mathrm{D}, b) = P$. Hence, his expected payoff is given by $f_i(\boldsymbol{a}) = \sum_\omega u_i(a_i, \omega_i)\sigma(\boldsymbol{\omega}|\boldsymbol{a})$. The expected payoff is determined by only action profile $\boldsymbol{a}$ regardless of signal profile $\boldsymbol{\omega}$. Thus, the expected payoff matrix is given by

$$\begin{array}{cc} & \begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array} \\ \begin{array}{c} \mathrm{C} \\ \mathrm{D} \end{array} & \begin{pmatrix} R_E & S_E \\ T_E & P_E \end{pmatrix} \end{array}. \tag{1}$$

According to $f_i(\boldsymbol{a})$, R_E, S_E, T_E, and P_E are derived as $R_E = R(1-\epsilon-\xi)+S(\epsilon+\xi)$, $S_E = S(1-\epsilon-\xi)+R(\epsilon+\xi)$, $T_E = T(1-\epsilon-\xi)+P(\epsilon+\xi)$, $P_E = P(1-\epsilon-\xi)+T(\epsilon+\xi)$, respectively. We assume that $T_E > R_E > P_E > S_E$ and $2R_E > T_E + S_E$, which dictate the RPD condition with observation errors.

In this paper, we introduce a discount factor to the RPD game with private monitoring. Although the game is played repeatedly over an infinite time horizon, the payoff will be discounted over rounds. Let w be the discount factor. Then, player i's discounted payoff of action profiles $\boldsymbol{a}^t, t \in \{0, 1, ..., \infty\}$ is $w^t f_i(\boldsymbol{a}^t)$ where t is a round. Finally, the average discounted payoff of player i is $s_i = (1-w)\sum_{t=0}^{\infty} w^t f_i(\boldsymbol{a}^t)$.

Consider each player i that adopts memory-one strategies with which they use only the outcome of the last round to decide the action to be submitted in the current round.

There are four types of outcomes, which are Cg, Cb, Dg and Db. Cg means the outcome when player i cooperated and observed the signal g. Cb means the outcome when player i cooperated and observed the signal b, and so forth. We define the conditional probability that player X cooperates when each outcome is realized as p_j, $j \in \{1,2,3,4\}$. Also, we define the probability that X cooperates in the first round as p_0. Thus, X's strategy is specified by $\boldsymbol{p} = (p_1, p_2, p_3, p_4; p_0)$. Similarly, Y's strategy is specified by $\boldsymbol{q} = (q_1, q_2, q_3, q_4; q_0)$.

Because both players adopt a memory-one strategy, the stochastic state of the two players is described by $\boldsymbol{v}(t) = (v_1(t), v_2(t), v_3(t), v_4(t))$, where the figures 1, 2, 3 and 4 of v mean the stochastic state (C,C), (C,D), (D,C) and (D,D), respectively. $v_1(t)$ is the probability that both players cooperate in round t, $v_2(t)$ is the probability that X cooperates and Y defects in round t, and so forth. We define the state transition matrix as M. The stochastic state of two players in round $t+1$ is calculated by $\boldsymbol{v}(t+1) = \boldsymbol{v}(t)M$. Then, the expected payoff to player X in round t is given by $\boldsymbol{v}(t)\boldsymbol{S}_X$, where $\boldsymbol{S}_X^T = (R_E, S_E, T_E, P_E)$. The expected per-round payoff to player X in the repeated game is given by

$$s_X = (1-w)\sum_{t=0}^{\infty} w^t \boldsymbol{v}(t)\boldsymbol{S}_X = \boldsymbol{v} \cdot \boldsymbol{S}_X \tag{2}$$

where I is the 4×4 identity matrix and the mean distribution $\boldsymbol{v}^T = (1-w)\boldsymbol{v}(0)(I - wM)^{-1}$ (Hilbe et al., 2015).

By conducting mathematical analyses, we found that Eq. (2) can be represented by a determinant form as follows even with observation errors and a discount factor, as Press and Dyson did without errors and no discount factor:

$$s_X = \boldsymbol{v} \cdot \boldsymbol{S}_X = \frac{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{S}_X)}{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{1})}, \tag{3}$$

where the determinant $D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{f})$ is represented by

$$\begin{vmatrix} \dots & w(\mu p_1 + \eta p_2) - 1 + p_0(1-w) & w(\mu q_1 + \eta q_2) - 1 + q_0(1-w) & f_1 \\ \dots & w(\eta p_1 + \mu p_2) - 1 + p_0(1-w) & w(\mu q_3 + \eta q_4) + q_0(1-w) & f_2 \\ \dots & w(\mu p_3 + \eta p_4) + p_0(1-w) & w(\eta q_1 + \mu q_2) - 1 + q_0(1-w) & f_3 \\ \dots & w(\eta p_3 + \mu p_4) + p_0(1-w) & w(\eta q_3 + \mu q_4) + q_0(1-w) & f_4 \end{vmatrix}, \tag{4}$$

where $\mu = 1 - \epsilon - \xi$, $\eta = \epsilon + \xi$, arbitrary vector $\boldsymbol{f} = (f_1, f_2, f_3, f_4)$ and the first column is omitted. Similarly, player Y's expected payoff s_Y is derived by replacing $\boldsymbol{S}_X$ with $\boldsymbol{S}_Y^T = (R_E, T_E, S_E, P_E)$ in Eq. (3).

Here, the linear combination of s_X and s_Y is given by

$$\alpha s_X + \beta s_Y + \gamma = \frac{D(\boldsymbol{p}, \boldsymbol{q}, \alpha \boldsymbol{S_X} + \beta \boldsymbol{S_Y} + \gamma \boldsymbol{1})}{D(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{1})}. \tag{5}$$

If the numerator of the right side of Eq. (5) is zero, that is, $D(\boldsymbol{p}, \boldsymbol{q}, \alpha \boldsymbol{S}_X + \beta \boldsymbol{S}_Y + \gamma \boldsymbol{1}) = 0$, the payoff relationship between s_X and s_Y becomes linear.

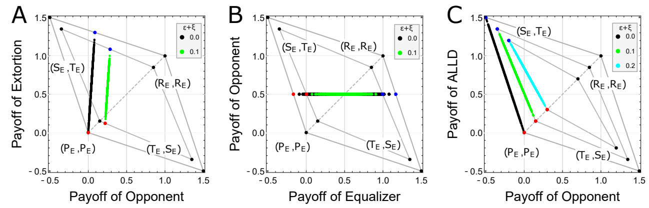

Figure 1: The payoff relationships between two players in the RPD game under observation errors when $w = 0.9$. Error rates $\epsilon + \xi$ are shown in the legends. (A) Extortioner (subclass of ZD), (B) Equalizer (subclass of ZD), and (C) ALLD (subclass of unconditional strategies) vs. 1,000 randomly generated strategies. $(T, R, P, S) = (1.5, 1, 0, -0.5)$.

Using the following determinant theorem: For an $n \times n$ matrix A, $\det(A) = 0 \Leftrightarrow$ The columns of matrix A are dependent vectors, we searched for all of X's strategies which satisfy Eq. (5). As a result, we found that the only strategies that impose a linear relationship between the two players' payoffs are either

$$\begin{aligned} w(\mu p_1 + \eta p_2) - 1 + p_0(1-w) &= \alpha R_E + \beta R_E + \gamma \\ w(\eta p_1 + \mu p_2) - 1 + p_0(1-w) &= \alpha S_E + \beta T_E + \gamma \\ w(\mu p_3 + \eta p_4) + p_0(1-w) &= \alpha T_E + \beta S_E + \gamma \\ w(\eta p_3 + \mu p_4) + p_0(1-w) &= \alpha P_E + \beta P_E + \gamma \end{aligned} \tag{6}$$

or

$$p_0 = p_1 = p_2 = p_3 = p_4. \tag{7}$$

The former corresponds to ZD strategies and the latter corresponds to unconditional strategies, respectively. Figure 1 shows some examples of ZD and unconditional strategies under observation errors where payoffs are discounted.

In conclusion, we derived the determinant form of the two player's expected payoff in the RPD game with a discount factor and observation errors. Then, we analytically revealed that the only strategy sets that enforce a linear payoff relationship are either the ZD strategies or the unconditional strategies in that situation. In repeated games, implementation errors (trembling hands) may be more likely to occur than observation errors. We can derive ZD strategies even under implementation errors because the same analysis used here can be applied. Thus, our technique is not limited to one particular errors but covers broad types of errors. This paper contributes to a deep understanding of ZD strategies in society. For further details, see our preprint (Mamiya and Ichinose, 2020).

References

Hao, D., Rong, Z., and Zhou, T. (2015). Extortion under uncertainty: Zero-determinant strategies in noisy games. *Phys. Rev. E*, 91:052803.

Hilbe, C., Traulsen, A., and Sigmund, K. (2015). Partners or rivals? Strategies for the iterated prisoner's dilemma. *Games Econ. Behav.*, 92:41–52.

Ichinose, G. and Masuda, N. (2018). Zero-determinant strategies in finitely repeated games. *J. Theor. Biol.*, 438:61–77.

Mamiya, A. and Ichinose, G. (2019). Strategies that enforce linear payoff relationships under observation errors in repeated prisoner's dilemma game. *J. Theor. Biol.*, 477:63–76.

Mamiya, A. and Ichinose, G. (2020). Zero-determinant strategies under observation errors in repeated games. *bioRxiv*.

Press, W. H. and Dyson, F. J. (2012). Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci. USA*, 109:10409–10413.