

ゼロ行列式戦略とその拡張

一ノ瀬 元喜, 間宮 安曇

ゼロ行列式戦略とは 2012 年に Press と Dyson が発見した繰り返しゲームにおける特徴的な戦略である。この戦略は、相手プレイヤーがどのような戦略をとっても相手と自身に線形の利得関係を強いる。この性質により、ゼロ行列式戦略は相手プレイヤーに対して利得の不公平な分配を強制することができる。本稿では、ゼロ行列式戦略とは何かをまず説明してから、その数理的な導出について記す。次に著者らが発展させてきたゼロ行列式戦略に関する割引因子と観測エラーの導入という拡張を紹介し、その解析によって得られた結果について記す。

キーワード：ゼロ行列式戦略, 囚人のジレンマゲーム, 繰り返しゲーム, 観測エラー, 割引因子

1. はじめに

ゲーム理論の繰り返しゲームにおいて、自分が協力的に振る舞うことによって後に相手からの協力を期待できるとする直接互惠 [1-3] の研究は長年行われてきたが、2012 年の W. H. Press と F. J. Dyson が米国科学アカデミー紀要 (*Proc. Natl. Acad. Sci. USA*) で発表した “Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent: 繰り返し囚人のジレンマは、あらゆる進化的な相手を支配する戦略を含む” というタイトルの論文 [4] は関連分野の研究者に大きな驚きをもって迎えられた。後に Zero-determinant strategy (著者らは日本語では、「ゼロ行列式戦略」と呼称している) と呼ばれることになったこの戦略は、相手がどのような戦略をとってきても、相手の 1 回のゲーム (ラウンド) あたりの期待利得をある一定の値に固定したり、相手には絶対負けしない (最悪でも引き分けて、それ以外は相手より必ず高い利得となる) 戦略を含む。これらの特徴は、ゼロ行列式戦略 (以降、ZD 戦略) が相手の戦略にかかわらずに、自分と相手との間に線形の利得関係を強制することができるという性質から来ている。ZD 戦略の発見に触発され、世界中の関連分野の研究者がさまざまな拡張を行ってきたが、著者らもここ数年、観測エラーや割引因子という現実的な要素を加えた ZD 戦略について理論を発展させてきた。本稿では、まず ZD 戦略とは何

かを説明する。その後、ZD 戦略の導出を行う。最後に、著者らが行ってきた ZD 戦略に関するさまざまな拡張について記す。

2. ZD 戦略とは何か

ZD 戦略とは、繰り返しゲームにおいて、相手がどのような戦略をとったとしても、相手に線形の利得関係を強いることができる戦略である [4]。図 1(A), (B) にその一例を示す。これは、プレイヤー X と Y の 2 人のプレイヤーが囚人のジレンマゲームを無限回繰り返したときの 1 回のゲーム (ラウンド) あたりの各プレイヤーの期待利得である s_X (縦軸) と s_Y (横軸) の関係を示している。

どちらの図でもプレイヤー X が ZD 戦略をとっている。図 1(A) の ZD 戦略は、Equalizer と呼ばれている ZD の特殊ケースの戦略であり、図 1(B) の ZD 戦略は Extortion (Extortioner とも呼称される) と呼ばれている別の ZD の特殊ケースの戦略である [4]。ZD 戦略について、図では、それぞれ一つの Equalizer 戦略と一つの Extortioner 戦略に固定しているが、プレイヤー Y については、ランダムに 1,000 戦略生成 + ALLC (無条件協力) + ALLD (無条件裏切り) の計 1,002 戦略用意している。ZD 戦略対ランダム戦略の期待利得を黒点、ZD 戦略対 ALLC (無条件協力) 戦略の期待利得を青点 (グラフ上部の点)、ZD 戦略対 ALLD (無条件裏切り) 戦略の期待利得を赤点 (グラフ下部の点) で表している。点が集中しているので、1 本の線のように見えるが、実際には 1,002 個の点が打たれている。これらの図から、数値的にはあるが、ZD 戦略は相手 (プレイヤー Y) がどんな戦略 (ここではランダム戦略と ALLC と ALLD) をとっても相手に直線の利得関係を強いていることがわかる。

一方で ZD 戦略ではないと同様の図はどうなる

いちのせ げんき

静岡大学大学院総合科学技術研究科工学専攻

〒432-8561 浜松市中区城北 3-5-1

ichinose.genki@shizuoka.ac.jp

まみや あづみ

三菱電機 (株) 名古屋製作所

〒461-8670 名古屋市中区東区矢田南五丁目 1 番 14 号

azumi_azumi0917@icloud.com

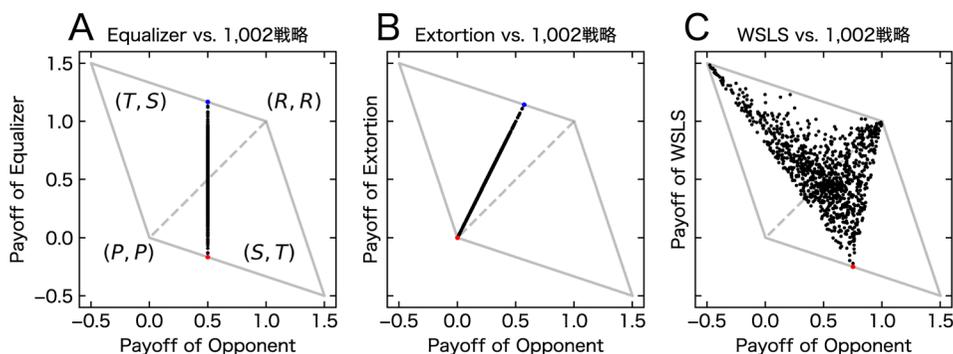


図 1 (A, B) ZD 戦略対 1,002 戦略 (ランダム + ALLC + ALLD) の期待利得の関係 (A) Equalizer 戦略 $\mathbf{p} = (2/3, 1/3, 2/3, 1/3)$. 相手の点 s_Y を常に $s_Y = 0.5$ に固定している. (B) Extortioneer 戦略 $\mathbf{p} = (0.9, 0.65, 0.25, 0)$ で相手の 2 倍の得点を得る ($s_X = 2s_Y$). (C) WSLS 戦略対 1,002 戦略の期待利得の関係. 利得ベクトルはすべて $(T, R, P, S) = (1.5, 1, 0, -0.5)$.

だろうか. 図 1(C) はよく知られた Win-Stay-Lose-Shift (WSLS) 戦略 [5] (Pavlov 戦略 [6] と呼ばれる) である. この図では縦軸が WSLS 戦略の期待値で横軸が相手の利得の期待値である. 相手の戦略はここでも同様に 1,002 戦略生成しており, WSLS 戦略と相手の一つの戦略の利得の期待値の組が一つの点で表されている. この図から, 一般的には, WSLS 戦略は ZD 戦略ではない (相手に直線の利得関係を強いていない) ことがわかる¹.

2.1 Equalizer 戦略と Extortioneer 戦略

ZD 戦略は相手に直線の利得関係を強いるが, 果たしてこれはどのような意味をもつのだろうか. これだけだとわかりにくい, 上述した Equalizer 戦略と Extortioneer 戦略はわかりやすい意味をもつ ZD 戦略の特殊ケースである. ここではこれらの戦略について詳しく説明する.

Equalizer と呼ばれる戦略はその名のとおりに, 相手がどんな戦略をとっても相手の利得をある固定した (equal の) 点数にさせることができる. 図 1(A) を見ると直線が横軸に対して垂直になっている. これはプレイヤー Y の期待利得 s_Y が固定した点数になることを意味する. 具体的に述べると, ここでは記憶 1 戦略 $\mathbf{p} = (2/3, 1/3, 2/3, 1/3)$ (記憶 1 戦略の定義は後述する) で表される Equalizer は, 繰り返しゲームの利得ベクトルが $(T, R, P, S) = (1.5, 1, 0, -0.5)$ (変数 T, R, P, S については後述する) となっているとき, 相手がどんな戦略をとっても相手の利得の期待値 s_Y を 0.5 に固定させることができる. つまり, Equalizer 戦

略は相手の利得の期待値をコントロールすることができる. その一方で, Equalizer 戦略の方は, 垂直線が縦に広がっていることからわかるように, 相手の戦略に依存してさまざまな利得の期待値になることがわかる.

次に Extortioneer 戦略について見ていく. Extortioneer の定義は, 「 (P, P) を始点として直線の傾きが 1 より大きい ZD 戦略」である. これが意味するのは, 相手がどんな戦略をとっても絶対に相手に負けないということである. ALLD などが相手の場合, (P, P) で引き分けとなってしまいが, この引き分けになる以外の相手には, どんな相手に対しても相手の利得よりも高くなる. 具体的に述べると, 図 1(B) は, ゲームの利得ベクトルが $(T, R, P, S) = (1.5, 1, 0, -0.5)$ において, $\mathbf{p} = (0.9, 0.65, 0.25, 0)$ という Extortioneer 戦略の一例を示している. このとき, 直線の傾きは 2 となる. $s_X = 2s_Y$ の直線となるので X の期待利得は Y の 2 倍となる. ここでは直線の傾きが 2 の例を示したが, 直線の傾きが 2 より小さいものや 2 よりも大きいものも作れ, 無限の Extortioneer 戦略を考えることができる. 以上のように ZD 戦略の中の Equalizer 戦略や Extortioneer 戦略は相手に対して利得の不公平な分配を一方向的に強いることができる.

3. ZD 戦略の導出

前節までで, グラフと言葉の説明を用いて ZD 戦略の特徴を記してきた. 本節では, ZD 戦略がどのようにして発見されたのかを導出する. ZD 戦略のゼロ行列の意味は, 行列式を 0 にするということである. Press と Dyson はプレイヤー X と Y の 2 人がゲーム

¹ 「一般的には」と書いたのは, 利得行列の値によっては WSLS 戦略も ZD 戦略となることがあるからである.

を無限回繰り返したときの1回のゲーム(ラウンド)あたりの利得の期待値が行列式の形で表せることを示した[4]。プレイヤーの期待利得がなぜ行列式の形で表せるか、またその行列式を0にするということがどういふことかを以下で順を追って説明する。

3.1 導出

2人のプレイヤー $i \in \{X, Y\}$ の繰り返し囚人のジレンマゲームを考える。それぞれのプレイヤーは、毎回のラウンドで、CまたはDのいずれかの行動を選択する。Cは協力、Dは裏切りを意味する。このゲームの利得行列は以下で与えられる。

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array} \quad (1)$$

この行列の要素はプレイヤー X が繰り返しゲームの1回のラウンドで得る利得を示している。行がプレイヤー X の行動、列がプレイヤー Y の行動に対応する。2人のプレイヤーがどちらもCを取ったとき、プレイヤー X は R の利得を得る。プレイヤー X がC、プレイヤー Y がDの行動をとったとき、プレイヤー X は S の利得を得る。プレイヤー X がD、プレイヤー Y がCの行動をとったとき、プレイヤー X は T の利得を得る。2人のプレイヤーがどちらもDを取ったとき、プレイヤー X は P の利得を得る。行をプレイヤー Y の行動、列をプレイヤー X の行動に変更すると、この行列はプレイヤー Y が得る利得を示す。ここで、ゲームが囚人のジレンマであるための条件 $T > R > P > S$ を仮定する。 $T > R$ かつ $P > S$ より、どちらのプレイヤーも、相手の戦略にかかわらず、CよりもDを取ることでより高い利得を得る。さらに、 $2R > T + S$ の条件を仮定する。この条件は、繰り返しのゲームにおいて、相互協力し続ける方が、一方がC、もう一方がDを交互に取るよりも利得が高くなることを保証する。

ここで2人のプレイヤーが記憶1戦略(Memory-one strategy)を採用している状況を考える。記憶1戦略とは、今回の戦略を決めるために1ラウンド前の結果(自分と相手の行動)の情報しか用いない戦略である。プレイヤー X の記憶1戦略を $\mathbf{p} = (p_1, p_2, p_3, p_4)$ で定義する。 p_1 は、前回のゲームで自分がC、相手がCであったとき、今回のゲームで自分がCを取る確率である。 p_2 は、前回のゲームで自分がC、相手がDであったとき、今回のゲームで自分がCを取る確率である。 p_3 は、前回のゲームで自分がD、相手がCであった

とき、今回のゲームで自分がCを取る確率である。 p_4 は、前回のゲームで自分がD、相手がDであったとき、今回のゲームで自分がCを取る確率である。同様にしてプレイヤー Y の記憶1戦略を $\mathbf{q} = (q_1, q_2, q_3, q_4)$ で定義する。 $q_j (j \in \{1, 2, 3, 4\})$ の意味は \mathbf{p} のときと同様である。

2人のプレイヤーとも記憶1戦略を採用すると仮定しているので、ラウンド $t (t \geq 0)$ での2人のプレイヤーの確率分布は、 $\mathbf{v}(t) = (v_1(t), v_2(t), v_3(t), v_4(t))$ と書ける。ここで、 v_1 は、ラウンド t で両者がCとなる確率、 v_2 は、ラウンド t でプレイヤー X がC、 Y がDとなる確率、 v_3 は、ラウンド t でプレイヤー X がD、 Y がCとなる確率、 v_4 は、ラウンド t で両者がDとなる確率である。次のラウンド $t+1$ での確率分布 $\mathbf{v}(t+1)$ は $\mathbf{v}(t)$ に以下で表されるゲームの遷移確率行列 M を掛けて $\mathbf{v}(t+1) = \mathbf{v}(t)M$ で求めることができる。

$$M = \begin{pmatrix} p_1 q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2 q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3 q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4 q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix} \quad (2)$$

この行列の行は、順に前回のゲームの状態(C, C), (C, D), (D, C), (D, D)を表す。同様に列は、順に次のゲームの状態(C, C), (C, D), (D, C), (D, D)を表す。したがって、たとえば M の1行2列は、ゲームの状態が(C, C)から(C, D)に移る遷移確率となる。

$\mathbf{v}(t+1)$ は $\mathbf{v}(t)$ にのみ依存するので、この状態遷移はマルコフ連鎖である。マルコフ連鎖がエルゴードの場合、そのマルコフ連鎖には次のような定常分布が存在する。

$$\mathbf{v} = \mathbf{v}M \quad (3)$$

ここで $M' \equiv M - I$ とすると式(3)は $\mathbf{v}M' = \mathbf{0}$ と変形できる。行列 M' にクラメルの公式を適用すると

$$\text{adj}(M')M' = O \quad (4)$$

が得られる。ここで $\text{adj}(M')$ は M' の余因子行列である。 $\mathbf{v}M' = \mathbf{0}$ かつ $\text{adj}(M')M' = O$ より、 $\text{adj}(M')$ のすべての行は \mathbf{v} に比例する。したがって、 \mathbf{v} と任意のベクトル \mathbf{f} の内積は、 $\text{adj}(M')$ の任意の行と \mathbf{f} の内積に比例する。たとえば、 $\text{adj}(M')$ の4行目を用いると $\mathbf{v} \cdot \mathbf{f} \propto C_{14}f_1 + C_{24}f_2 + C_{34}f_3 + C_{44}f_4$ となる。ここで C_{ij} は $\text{adj}(M')$ の i 列 j 列の要素を意味する。これは行列 M' の4列目を \mathbf{f} に置き換えた行列式の余因子展開に等しいため、次のように書き換えることができる。

$$\mathbf{v} \cdot \mathbf{f} \propto \begin{vmatrix} p_1 q_1 - 1 & p_1(1 - q_1) & (1 - p_1)q_1 & f_1 \\ p_2 q_3 & p_2(1 - q_3) - 1 & (1 - p_2)q_3 & f_2 \\ p_3 q_2 & p_3(1 - q_2) & (1 - p_3)q_2 - 1 & f_3 \\ p_4 q_4 & p_4(1 - q_4) & (1 - p_4)q_4 & f_4 \end{vmatrix} \quad (5)$$

2 列目と 3 列目に最初の列を足しても行列式は変わらないので、これはさらに、次のように書き換えることができる。

$$\mathbf{v} \cdot \mathbf{f} \propto \begin{vmatrix} p_1 q_1 - 1 & p_1 - 1 & q_1 - 1 & f_1 \\ p_2 q_3 & p_2 - 1 & q_3 & f_2 \\ p_3 q_2 & p_3 & q_2 - 1 & f_3 \\ p_4 q_4 & p_4 & q_4 & f_4 \end{vmatrix} \equiv D(\mathbf{p}, \mathbf{q}, \mathbf{f}) \quad (6)$$

式 (6) の行列式を正規化するために $D(\mathbf{p}, \mathbf{q}, \mathbf{1})$ で割ると、 $\mathbf{v} \cdot \mathbf{f}$ が得られる。

$$\mathbf{v} \cdot \mathbf{f} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{f})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})} \quad (7)$$

ここで $\mathbf{1} = (1, 1, 1, 1)$ である。任意のベクトル \mathbf{f} をプレイヤー X の利得ベクトル $\mathbf{S}_X = (R, S, T, P)$ で置き換えると

$$s_X = \mathbf{v} \cdot \mathbf{S}_X = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S}_X)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})} \quad (8)$$

が得られる。同様に \mathbf{f} をプレイヤー Y の利得ベクトル $\mathbf{S}_Y = (R, T, S, P)$ で置き換えると、

$$s_Y = \mathbf{v} \cdot \mathbf{S}_Y = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S}_Y)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})} \quad (9)$$

が得られる。 s_X や s_Y は、定常分布に利得ベクトルを掛けたものなので、それぞれのプレイヤーのラウンドあたりの期待利得である。以上の線形代数的な操作によって期待利得 s_X や s_Y を行列式の形で表すことができた。さらに、これらを線形結合したものは、 α, β, γ を実数とすると、

$$\alpha s_X + \beta s_Y + \gamma = \frac{D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})} \quad (10)$$

と、これも行列式の形で表すことができる。式 (10) の右辺の分子 $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})$ の値が 0 ならば、

$$\alpha s_X + \beta s_Y + \gamma = 0 \quad (11)$$

となり、2 人のプレイヤーの期待利得が線形の関係になる。ここで、式 (10) 右辺の分子の中身は以下のようになっている。

$$D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}) = \begin{vmatrix} p_1 q_1 - 1 & p_1 - 1 & q_1 - 1 & \alpha R + \beta R + \gamma \\ p_2 q_3 & p_2 - 1 & q_3 & \alpha S + \beta T + \gamma \\ p_3 q_2 & p_3 & q_2 - 1 & \alpha T + \beta S + \gamma \\ p_4 q_4 & p_4 & q_4 & \alpha P + \beta P + \gamma \end{vmatrix} \quad (12)$$

式 (12) の右辺の行列式の 2 列目に注目するとプレイヤー X の戦略 p_j ($j \in \{1, 2, 3, 4\}$) のみから構成される。行列式の性質から、もし 2 列目と 4 列目が等しければ行列式は 0 となるので、式 (10) の右辺も 0 となり、式 (11) が成り立つ。したがって、もしプレイヤー X が行列式の 4 列目と等しくなるように自身の戦略 p_j ($j \in \{1, 2, 3, 4\}$) を調整すれば、プレイヤー Y の戦略に関係なく (q_j ($j \in \{1, 2, 3, 4\}$) は入っていないので)、線形の利得関係を強いることができる。このように式 (10) の右辺の行列式を 0 にする戦略が ZD 戦略であり、この戦略は式 (12) の行列式の 2 列目と 4 列目が等しいという条件から以下の式で表される。

$$\begin{aligned} p_1 &= \alpha R + \beta R + \gamma + 1 \\ p_2 &= \alpha S + \beta T + \gamma + 1 \\ p_3 &= \alpha T + \beta S + \gamma \\ p_4 &= \alpha P + \beta P + \gamma \end{aligned} \quad (13)$$

3.2 Equalizer 戦略と Extortioner 戦略の表式

式 (13) において、 $\alpha = 0$ としたものが 2.1 節前半で説明した Equalizer 戦略である。 $\alpha = 0$ を式 (13) に代入することで以下の Equalizer 戦略の表式が得られる。

$$\begin{aligned} p_1 &= \beta R + \gamma + 1 \\ p_2 &= \beta T + \gamma + 1 \\ p_3 &= \beta S + \gamma \\ p_4 &= \beta P + \gamma \end{aligned} \quad (14)$$

この四つの方程式から β, γ を消去し、 p_1 と p_4 を用いて p_2, p_3 について解くと以下が得られる。

$$p_2 = \frac{p_1(T - P) - (1 + p_4)(T - R)}{R - P} \quad (15)$$

$$p_3 = \frac{(1 - p_1)(P - S) + p_4(R - S)}{R - P} \quad (16)$$

また式 (14) の連立方程式から p_2, p_3 を消去し、 β, γ を p_1 と p_4 で表したものと $\alpha = 0$ を式 (11) に代入して解くとプレイヤー Y の期待利得 s_Y の式が得られる。

$$s_Y = \frac{(1 - p_1)P + p_4R}{(1 - p_1) + p_4} \quad (17)$$

なお、式 (17) において、 $0 \leq p_1 \leq 1$ と $0 \leq p_4 \leq 1$ より ZD 戦略は $P \leq s_Y \leq R$ の任意の値を相手プレイヤーに強いることができることもわかっている [4].

ここで、図 1(A) を改めて見ると、 $\mathbf{p} = (2/3, 1/3, 2/3, 1/3)$ の Equalizer は相手プレイヤーの期待利得を $s_Y = 0.5$ に固定しているが、これは利得ベクトル $(T, R, P, S) = (1.5, 1, 0, -0.5)$ が与えられたときに $p_1 = 2/3, p_4 = 1/3$ と決めて式 (15)–(17) から p_2, p_3, s_Y について求めたものであることがわかる。

一方で、式 (13) において、 $\alpha = \phi (\neq 0), \beta = -\phi\chi, \gamma = \phi(\chi - 1)P$ としたものが 2.1 節後半で説明した Extortioner 戦略であり、以下の式で表される。

$$\begin{aligned} p_1 &= 1 + \phi[(R - P) - \chi(R - P)] \\ p_2 &= 1 + \phi[(S - P) - \chi(T - P)] \\ p_3 &= \phi[(T - P) - \chi(S - P)] \\ p_4 &= 0 \end{aligned} \quad (18)$$

プレイヤー X がこの Extortioner 戦略をとると、

$$s_X - P = \chi(s_Y - P) \quad (19)$$

の線形の利得関係を強いることができる。 χ は直線の傾きを示しており、図 1(B) を改めて見ると、これは利得ベクトル $(T, R, P, S) = (1.5, 1, 0, -0.5)$ が与えられたときに、式 (18) において、 $\phi = 0.1, \chi = 2$ と決めることで得られた $\mathbf{p} = (0.9, 0.65, 0.25, 0)$ の Extortioner 戦略だということがわかる。

4. ZD 戦略の拡張

この ZD 戦略について、著者らはさまざまな拡張を行ってきた [7–10]。ここでは、それらの拡張について紹介する。

4.1 割引因子の導入

繰り返しゲームにおいて、将来の利得は現在の価値よりも割引いて考え、プレイヤーは行動選択を行うことが自然である。したがって、利得に割引因子を導入した繰り返しゲームのモデルが伝統的に考えられてきた。割引因子はプレイヤーが将来与えられる利得を重視するパラメーターである。割引因子が大きいほど、将来の利得を重視し、小さいほど現在の利得を重視することを表す。次の期で得られる利得は現在の利得に δ ($0 < \delta < 1$) をかけることで、価値が割引かれる。割引因子付きゲームは、ゲームの終了期が決まっていない有限回ゲームとも解釈でき、このように解釈した場合、 δ はゲームが続く確率と解釈できる。ZD 戦略の利得の直線関係はプレイヤーが得る利得に依存する。

したがって、ZD 戦略に関して、将来の利得における割引因子の影響を調査することは重要であり、著者らはその影響を調査した [7]。ここでは、その概略と見つかった割引因子ありでの ZD 戦略について説明する。

著者らは、まず式 (11) から分析を開始した。目的は、線形の利得関係を課すプレイヤー X の記憶 1 戦略 \mathbf{p} を見つけることである。仮に \mathbf{p} が ZD 戦略だとすれば、プレイヤー Y のいくつかの代表的な戦略 \mathbf{q} にかかわらず、2 人の利得関係はある直線上に乗らなければならない。プレイヤー Y の代表的な戦略とは、たとえば $\mathbf{q} = (1, 1, 1, 1), \mathbf{q} = (0, 0, 0, 0), \mathbf{q} = (1, 0, 1, 0)$ など考えることができる。著者らは、割引因子ありでの s_X と s_Y の式を導いてから、これらを式 (11) に代入した。これにより、 \mathbf{q} を含まないいくつかの関係式を得ることができる。少なくともこれらの \mathbf{q} に対して、2 人のプレイヤーの期待利得は線形になっていなければならないので、これは必要条件を与える。このようにして得られた \mathbf{p} が十分条件も満たすかどうか確かめるため、この \mathbf{p} と任意の $\mathbf{q} = (q_1, q_2, q_3, q_4)$ を逆に式 (11) に代入した結果、 \mathbf{q} は消去され、この線形の関係式も満たされることがわかった。

この結果得られた、割引因子ありでの ZD 戦略の式は以下となった。

$$\begin{aligned} \delta p_1 - 1 + p_0(1 - \delta) &= \alpha R + \beta R + \gamma \\ \delta p_2 - 1 + p_0(1 - \delta) &= \alpha S + \beta T + \gamma \\ \delta p_3 + p_0(1 - \delta) &= \alpha T + \beta S + \gamma \\ \delta p_4 + p_0(1 - \delta) &= \alpha P + \beta P + \gamma \end{aligned} \quad (20)$$

この式は、割引因子を考えない、つまり $\delta = 1$ のときは、式 (13) と一致することがわかる。式 (20) において、 p_0 はラウンド $t = 0$ での C を取る確率であり、割引因子を考えたときは、この p_0 も ZD 戦略の式に入ることがわかった。

また分析の結果、ZD 戦略だけではなく、

$$p_0 = p_1 = p_2 = p_3 = p_4 \quad (21)$$

という unconditional strategy (無条件戦略) も相手がどんな戦略をとっても相手に線形の利得関係を強いることができる戦略であることも判明した [7, 11].

さらには、Equalizer 戦略や Extortioner 戦略が存在するための最小の割引因子 δ_c が以下であることを明らかにした。

$$\delta_c \equiv \max\left(\frac{T - R}{T - P}, \frac{P - S}{R - S}\right) \quad (22)$$

Equalizer 戦略は $\delta \geq \delta_c$ で存在し、Extortioner 戦略は $\delta > \delta_c$ で存在する。また Extortioner 戦略が存在するための最小の直線の傾き χ_c が以下であることも明らかにした。

$$\chi_c \equiv \max \left(\frac{R - S - \delta(P - S)}{\delta(T - P) - (T - R)}, \frac{T - P - \delta(T - R)}{\delta(R - S) - (P - S)} \right) \quad (23)$$

4.2 観測エラーの導入

現実社会では、ほかのプレイヤーの行動が完全に観測できない場合（不完全観測：imperfect monitoring）がある。相手の行動が直接観測できないとき、相手は何をやったか全くわからないわけではなく、多くの場合、プレイヤーたちの行動を不完全ながら表す何かのシグナルが観測できることが多い。たとえば、相手の行動に関するシグナル、「シグナル g (good)」あるいは「シグナル b (bad)」のどちらかが私的に観察されるとする。このシグナルには、ノイズがついており、相手が協力（または、裏切り）の場合、g（または、b）が観測されるが、周りの環境などの要因により、ある確率で逆の b（または、g）の信号を観測してしまう。このように、本来観測すべき信号ではない、逆の信号を観測してしまうことを観測エラーという。著者らの論文 [8] では、観測エラーを繰り返しゲームに導入し、私的不完全観測下での ZD 戦略について調査した。ここでは、その概略と見つかった観測エラーありでの ZD 戦略について説明する。

この論文ではまず、Hao et al. [12] が先にやったのと同様に、観測エラーありを含んだ形での式 (10) の $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})$ を式 (24) のように導いた。

ここで p, q について、前回のゲーム結果が Cg, Cb, Dg, Db のときの添字を、それぞれ 1, 2, 3, 4 とする。また ϵ はどちらか一方のプレイヤーに観測エラーが起こる確率であり、 ξ は両方のプレイヤーに観測エラーが起こる確率である。また $\tau = 1 - 2\epsilon - \xi$, $\mu = 1 - \epsilon - \xi$, $\eta = \epsilon + \xi$ である。その後、 \mathbf{q} に関係なく、式 (11) を成立させる $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}) = 0$ となる戦略 \mathbf{p} を以下の定理を用いてすべて探索した。

$$D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}) = \begin{vmatrix} \tau p_1 q_1 + \epsilon p_1 q_2 + \epsilon p_2 q_1 + \xi p_2 q_2 - 1 & \mu p_1 + \eta p_2 - 1 & \mu q_1 + \eta q_2 - 1 & \alpha R_E + \beta R_E + \gamma \\ \epsilon p_1 q_3 + \xi p_1 q_4 + \tau p_2 q_3 + \epsilon p_2 q_4 & \eta p_1 + \mu p_2 - 1 & \mu q_3 + \eta q_4 & \alpha S_E + \beta T_E + \gamma \\ \epsilon p_3 q_1 + \tau p_3 q_2 + \xi p_4 q_1 + \epsilon p_4 q_2 & \mu p_3 + \eta p_4 & \eta q_1 + \mu q_2 - 1 & \alpha T_E + \beta S_E + \gamma \\ \xi p_3 q_3 + \epsilon p_3 q_4 + \epsilon p_4 q_3 + \tau p_4 q_4 & \eta p_3 + \mu p_4 & \eta q_3 + \mu q_4 & \alpha P_E + \beta P_E + \gamma \end{vmatrix} \quad (24)$$

定理 1. n 次正方行列 A に対して、

$$\det(A) = 0 \Leftrightarrow A \text{ の列ベクトルは 1 次従属}$$

が成り立つ。

その結果、観測エラーありでの ZD 戦略の式は以下となった。これは Hao et al. [12] が先に見つけたものと一致する。

$$\begin{aligned} \mu p_1 + \eta p_2 - 1 &= \alpha R_E + \beta R_E + \gamma \\ \eta p_1 + \mu p_2 - 1 &= \alpha S_E + \beta T_E + \gamma \\ \mu p_3 + \eta p_4 &= \alpha T_E + \beta S_E + \gamma \\ \eta p_3 + \mu p_4 &= \alpha P_E + \beta P_E + \gamma \end{aligned} \quad (25)$$

観測エラーが入るので、利得について $R_E = R(1 - \epsilon - \xi) + S(\epsilon + \xi)$, $S_E = S(1 - \epsilon - \xi) + R(\epsilon + \xi)$, $T_E = T(1 - \epsilon - \xi) + P(\epsilon + \xi)$, $P_E = P(1 - \epsilon - \xi) + T(\epsilon + \xi)$ の変換が必要である。式 (25) は、観測エラーがないとき ($\epsilon = 0, \xi = 0$) は、式 (13) と一致することがわかる。また割引因子ありの場合と同様に、観測エラーありの状況でも $p_1 = p_2 = p_3 = p_4$ で表される無条件戦略もまた線形の利得関係を強いる戦略であることも判明した。この無条件戦略は、式 (21) に対応するが、割引因子がない今回の場合は p_0 は含まれない (p_0 は何でもよい)。

4.3 割引因子と観測エラーの同時導入

次の論文 [9] では、割引因子と観測エラーを同時に考えた。前節と同様のやり方で数理解析を行った。まず割引因子と観測エラーが入ったうえでの式 (10) の $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})$ を求めた。式が長いのでここでは省略するが、これは、式 (24) にさらに割引因子 δ が入った複雑な形をしていた。その後、 \mathbf{q} に関係なく、この $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}) = 0$ を成立させる戦略 \mathbf{p} を定理 1 を用いてすべて探索した。

その結果、割引因子と観測エラーが両方ともある場合の ZD 戦略の式は以下となった。

$$\begin{aligned}
\delta(\mu p_1 + \eta p_2) - 1 + p_0(1 - \delta) &= \alpha R_E + \beta R_E + \gamma \\
\delta(\eta p_1 + \mu p_2) - 1 + p_0(1 - \delta) &= \alpha S_E + \beta T_E + \gamma \\
\delta(\mu p_3 + \eta p_4) + p_0(1 - \delta) &= \alpha T_E + \beta S_E + \gamma \\
\delta(\eta p_3 + \mu p_4) + p_0(1 - \delta) &= \alpha P_E + \beta P_E + \gamma
\end{aligned} \tag{26}$$

式 (26) は、割引因子も観測エラーもない ($\delta = 1, \epsilon = 0, \xi = 0$) ときは、式 (13) と一致することがわかる。割引因子 ($0 < \delta < 1$) のみある場合、式 (20) と一致することがわかる。観測エラーのみある ($\epsilon > 0, \xi > 0$) 場合、式 (25) と一致することがわかる。またこれまでの場合と同様に、割引因子と観測エラーが両方ある状況でも式 (21) で表される無条件戦略もまた線形の利得関係を強いる戦略であることも明らかにした。

図 2 はプレイヤー X はある固定した戦略、プレイヤー Y は図 1 と同様に 1,002 戦略ずつ生成した際の両者の期待利得の関係を図示したものである。図 2(A)–(D) は割引因子がない ($\delta = 1$) 場合、(E)–(H) が割引因子が $\delta = 0.9$ の場合である。すべての図において黒点が観測エラーがない場合、黄緑点が観測エラーが $\epsilon + \xi = 0.1$ の場合、水色点が $\epsilon + \xi = 0.2$ の場合である。(A) と (E) のプレイヤー X は WLSL であり、これは一般には ZD 戦略ではないので、線形の利得関係とならない。(B) と (F) はプレイヤー X が Extortioneer 戦略の場合である。数値的にはあるが、観測エラーがあっても (B)、観測エラーと割引因子の両方があっても (F)、線形の利得関係を強いていることがわかる²。(C) と (G) はプレイヤー X が Equalizer 戦略の場合である。これも数値的にはあるが、観測エラーがあっても (C)、観測エラーと割引因子の両方があっても (G)、プレイヤー Y の戦略にかかわらず s_Y は常に一定を強制されることがわかる。(D) と (H) はプレイヤー X が無条件戦略の一つの例である ALLD 戦略の場合である。著者らの論文で明らかになったように無条件戦略も線形の利得関係を相手に強いていることがわかる。

4.4 ZD 戦略の存在条件

図 2 で割引因子 δ や観測エラー ϵ, ξ があっても ZD 戦略が存在することはわかるが、それらがどのような値でも ZD 戦略は存在できるわけではなくそれぞれに閾値が存在する。したがって、著者らは ZD 戦略が存

² 観測エラーがない場合、Extortioneer 戦略は絶対負けがないが、エラーがあると、相手より点数が低くなってしまふ場合が存在することも知られている。実際、 (P_E, P_E) 付近では相手よりも点数が低い。エラーがある場合の Extortioneer 戦略は通常のものとは区別して Contingent Extortioneer と呼ばれている [8, 9, 12]。

在できる条件を明らかにすることに取り組んだ [10]。数理解析の詳細は省略するが、 $\alpha = 0$ の Equalizer 戦略が存在する条件 δ_c, η_c は以下のようになることを明らかにした。

$$\delta \geq \delta_c \equiv \max\left(\frac{T_E - R_E}{\mu(R_E - P_E) - \eta(T_E - S_E) + T_E - R_E}, \frac{P_E - S_E}{\mu(R_E - P_E) - \eta(T_E - S_E) + P_E - S_E}\right) \tag{27}$$

$$\eta < \eta_c \equiv \frac{R_E - P_E}{T_E + R_E - P_E - S_E} \tag{28}$$

一方で、Extortioneer 戦略を含む $\chi > 1$ の pcZD 戦略 (positively correlated ZD) が存在する条件は、 $\delta > \delta_c, \eta < \eta_c$ に加えて、以下の κ の範囲と χ_c なることを明らかにした。

$$\begin{aligned}
P_E + \frac{\eta}{\mu - \eta} \left(\frac{T_E - S_E}{\chi - 1} + P_E - S_E \right) &\leq \kappa \\
&\leq R_E - \frac{\eta}{\mu - \eta} \left(\frac{T_E - S_E}{\chi - 1} + T_E - R_E \right)
\end{aligned} \tag{29}$$

$$\chi \geq \chi_c \equiv \max(\chi_1, \chi_2) \tag{30}$$

ここで、

$$\chi_1 = 1 + \frac{(1 - \delta + 2\delta\eta)(T_E - S_E)}{\delta(\mu(R_E - P_E) - \eta(T_E - S_E)) - (1 - \delta)(T_E - R_E)} \tag{31}$$

$$\chi_2 = 1 + \frac{(1 - \delta + 2\delta\eta)(T_E - S_E)}{\delta(\mu(R_E - P_E) - \eta(T_E - S_E)) - (1 - \delta)(P_E - S_E)} \tag{32}$$

である。なお、式 (27) と式 (30) について、観測エラーがないときは式 (22) と式 (23) にそれぞれ一致することもわかる。

5. おわりに

本稿では、繰返しゲームにおいて Press と Dyson によって発見された ZD 戦略について、その紹介と著者らが行ってきた拡張について紹介した。まず、ZD 戦略は相手がどのような戦略をとっても自身と相手に線形の利得関係を強いることを数値計算の例で説明した。その後、ZD 戦略の導出を行った。プレイヤーの 1 ラウンドあたりの期待利得が行列式の形で表せることを説明し、そこから 2 人の利得の線形和も行列式の形になることを示した。ZD 戦略とは、この利得の線形和の行列式を 0 にするような戦略のことであり、そ

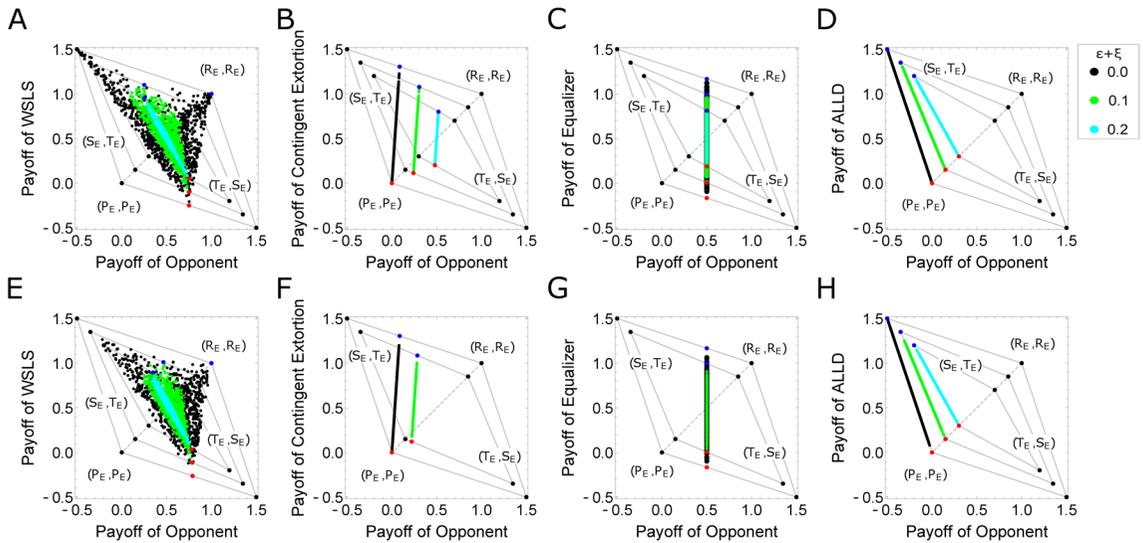


図2 繰り返し四人のジレンマゲームでの2人のプレイヤー間の期待利得の関係
 (A)–(D) : 割引因子がない ($\delta = 1$) の場合, (E)–(H) : 割引因子が $\delta = 0.9$ の場合,
 黒点: 観測エラーなし, 黄緑点: 観測エラーが $\epsilon + \xi = 0.1$, 水色点: $\epsilon + \xi = 0.2$.
 利得ベクトル: $(T, R, P, S) = (1.5, 1, 0, -0.5)$. (A, E) WSLs 戦略 vs. 1,002 戦
 略. (B, F) Extortioner 戦略 vs. 1,002 戦略. (C, G) Equalizer 戦略 vs. 1,002 戦
 略. (D, H) ALLD 戦略 vs. 1,002 戦略.

の式も示した。ZD 戦略の研究の発展として、著者らは割引因子や観測エラーなどの現実的な要素をモデルに組み込み、このような状況でも ZD 戦略が存在することを示した。数理解析の中で、ZD 戦略だけでなく無条件戦略も線形の利得関係を強いることを明らかにした。さらには、ZD 戦略が存在できる割引因子、観測エラーのパラメーターの下限についても解析的に求めた。

ZD 戦略については現在でもさまざまな拡張が行われている。ZD 戦略は 2 人プレイヤーで四人のジレンマゲームだけに存在するものではなく、3 人以上のプレイヤー [13–16] やほかのゲーム [4, 17, 18] でも存在することが知られている。一方で、進化ゲームの研究者により、ZD 戦略の進化についてさまざまな研究が行われてきている [11, 14, 19–28]。その中で、Extortioner 戦略は対一のゲームでは強いが、集団を考え、戦略が進化する状況を考えて場合には、強い戦略ではないこと、むしろ ZD 戦略のほかの特殊ケースである Generous 戦略が進化的に強いことなどが既に明らかとなっている [19]。ZD 戦略を含めて“Partner と Rival”あるいは“Friendly-rivalry”といった形で戦略の分類を行うことによって [29–32]、それらの戦略がもつ特徴も徐々に明らかとなってきており、今後、繰り返しゲームにおける戦略のさらなる理解が進むであろう。

著者らもさらに ZD 戦略に関する研究を進めている。

$\chi > 1$ の pcZD 戦略は相手に線形の利得関係を強いることができるだけでなく、利得を改善したいと考える適応的な相手プレイヤーを最終的に ALLC (無条件協力) に導くことができるという性質がある。この性質が割引因子ありの状況でも成り立つか現在解析中であり、研究成果がまとまり次第、発表したいと考えている。

謝辞 本稿を執筆する機会を作っていただいた朝日弓未先生、吉村仁先生、伊東啓先生に心よりお礼を申し上げます。

参考文献

- [1] R. L. Trivers, “The evolution of reciprocal altruism,” *Quarterly Review of Biology*, **46**, pp. 35–57, 1971.
- [2] M. A. Nowak, *Evolutionary Dynamics*, Harvard University Press, 2006.
- [3] K. Sigmund, *The Calculus of Selfishness*, Princeton University Press, 2010.
- [4] W. H. Press and F. J. Dyson, “Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent,” *Proceedings of the National Academy of Sciences of the United States of America*, **109**, pp. 10409–10413, 2012.
- [5] M. Nowak and K. Sigmund, “A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner’s dilemma game,” *Nature*, **364**, pp. 56–58, 1993.
- [6] D. Kraines and V. Kraines, “Learning to cooperate with Pavlov for the iterated Prisoner’s Dilemma with noise,” *Theory and Decision*, **35**,

pp. 107–150, 1993.

- [7] G. Ichinose and N. Masuda, “Zero-determinant strategies in finitely repeated games,” *Journal of Theoretical Biology*, **438**, pp. 61–77, 2018.
- [8] A. Mamiya and G. Ichinose, “Strategies that enforce linear payoff relationships under observation errors in repeated prisoner’s dilemma game,” *Journal of Theoretical Biology*, **477**, pp. 63–76, 2019.
- [9] A. Mamiya and G. Ichinose, “Zero-determinant strategies under observation errors in repeated games,” *Physical Review E*, **102**, 032115, 2020.
- [10] A. Mamiya and G. Ichinose, “Conditions for the existence of zero-determinant strategies under observation errors in repeated games,” *Journal of Theoretical Biology*, **526**, 110810, 2021.
- [11] C. Hilbe, M. A. Nowak and A. Traulsen, “Adaptive dynamics of extortion and compliance,” *PLoS ONE*, **8**, e77886, 2013.
- [12] D. Hao, Z. Rong and T. Zhou, “Extortion under uncertainty: Zero-determinant strategies in noisy games,” *Physical Review E*, **91**, 052803, 2015.
- [13] M. Ueda and T. Tanaka, “Linear algebraic structure of zero-determinant strategies in repeated games,” *PLoS ONE*, **15**, e0230973, 2020.
- [14] C. Hilbe, B. Wu, A. Traulsen and M. A. Nowak, “Evolutionary performance of zero-determinant strategies in multiplayer games,” *Journal of Theoretical Biology*, **374**, pp. 115–124, 2015.
- [15] C. Hilbe, B. Wu, A. Traulsen and M. A. Nowak, “Cooperation and control in multiplayer social dilemmas,” *Proceedings of the National Academy of Sciences of the United States of America*, **111**, pp. 16425–16430, 2014.
- [16] L. Pan, D. Hao, Z. Rong and T. Zhou, “Zero-determinant strategies in iterated public goods game,” *Scientific Reports*, **5**, 13096, 2015.
- [17] M. A. Taha and A. Ghoneim, “Zero-determinant strategies in repeated asymmetric games,” *Applied Mathematics and Computation*, **369**, 124862, 2019.
- [18] A. McAvooy and C. Hauert, “Autocratic strategies for iterated games with arbitrary action spaces,” *Proceedings of the National Academy of Sciences of the United States of America*, **113**, pp. 3573–3578, 2016.
- [19] A. J. Stewart and J. B. Plotkin, “From extortion to generosity, evolution in the Iterated Prisoner’s Dilemma,” *Proceedings of the National Academy of Sciences of the United States of America*, **110**, pp. 15348–15353, 2013.
- [20] C. Adami and A. Hintze, “Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything,” *Nature Communications*, **4**, 2193, 2013.
- [21] C. Hilbe, M. A. Nowak and K. Sigmund, “Evolution of extortion in iterated prisoner’s dilemma games,” *Proceedings of the National Academy of Sciences of the United States of America*, **110**, pp. 6913–6918, 2013.
- [22] A. Szolnoki and M. Perc, “Evolution of extortion in structured populations,” *Physical Review E*, **89**, 022804, 2014.
- [23] Z. X. Wu and Z. Rong, “Boosting cooperation by involving extortion in spatial prisoner’s dilemma games,” *Physical Review E*, **90**, 062102, 2014.
- [24] J. Liu, Y. Li, C. Xu and P. M. Hui, “Evolutionary behavior of generalized zero-determinant strategies in iterated prisoner’s dilemma,” *Physica A*, **430**, pp. 81–92, 2015.
- [25] X. Xu, Z. Rong, Z. X. Wu, T. Zhou and C. K. Tse, “Extortion provides alternative routes to the evolution of cooperation in structured populations,” *Physical Review E*, **95**, 052302, 2017.
- [26] J. Wang and J. Guo, “A synergy of punishment and extortion in cooperation dilemmas driven by the leader,” *Chaos Solitons Fractals*, **119**, pp. 263–268, 2019.
- [27] Y. Mao, X. Xu, Z. Rong and Z. X. Wu, “The emergence of cooperation-extortion alliance on scale-free networks with normalized payoff,” *EPL*, **122**, 50005, 2018.
- [28] X. Xu, Z. Rong, Z. Tian and Z. X. Wu, “Timescale diversity facilitates the emergence of cooperation-extortion alliances in networked systems,” *Neurocomputing*, **350**, pp. 195–201, 2019.
- [29] C. Hilbe, A. Traulsen and K. Sigmund, “Partners or rivals? Strategies for the iterated prisoner’s dilemma,” *Games and Economic Behavior*, **92**, pp. 41–52, 2015.
- [30] C. Hilbe, K. Chatterjee and M. A. Nowak, “Partners and rivals in direct reciprocity,” *Nature Human Behaviour*, **2**, pp. 469–477, 2018.
- [31] Y. Murase and S. K. Baek, “Five rules for friendly rivalry in direct reciprocity,” *Scientific Reports*, **10**, 16904, 2020.
- [32] Y. Murase and S. K. Baek, “Friendly-rivalry solution to the iterated n-person public-goods game,” *PLoS Computational Biology*, **17**, e1008217, 2021.